AN EXPLORATION OF HOW THE MILITARY INTELLIGENCE COMMUNITY MAY MITIGATE UNINTENDED BIAS WITHIN MACHINE LEARNING SYSTEMS

A thesis presented to the Faculty of the U.S. Army Command and General Staff College in partial fulfillment of the requirements for the degree

MASTER OF MILITARY ART AND SCIENCE
Information Advantage Scholars

by

BETHANY BASHOR, MAJOR, U.S. ARMY
MGIS, Pennsylvania State University, State College, PA, 2020

Fort Leavenworth, Kansas
2022

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED *(From - To)*
10-06-2022	Master's Thesis	AUG 2021 – JUN 2022

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
An Exploration of How the Military Intelligence Community May Mitigate Unintended Bias within Machine Learning Systems	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
Bethany Bashor	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Command and General Staff College ATTN: ATZL-SWD-GD Fort Leavenworth, KS 66027-2301	8. PERFORMING ORG REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution is Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

As the military modernizes with Artificial Intelligence/Machine Learning systems, the personnel in the military intelligence community will need to fully understand and integrate these systems into the Joint Intelligence Process (JIP). To guide the force in the development and integration of this technology, the Department of Defense (DOD) publicly released the "DOD Adopts Ethical Principles for Artificial Intelligence," which included "equitable" as a principle to mitigate "unintended bias." While there are multiple technical best practices to mitigate unintended bias, there are also nontechnical best practices that the military intelligence community could adopt from commercial industry. This qualitative, multiple case study and cross-case synthesis explores these commercial industry nontechnical best practices to mitigate unintended bias in the JIP.

15. SUBJECT TERMS
Department of Defense; Military Intelligence; Machine Learning; Unintended Bias

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. PHONE NUMBER *(include area code)*
(U)	(U)	(U)	(U)	113	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

MASTER OF MILITARY ART AND SCIENCE

THESIS APPROVAL PAGE

Name of Candidate: Bethany D. Bashor

Thesis Title: An Exploration of How the Military Intelligence Community May Mitigate Unintended Bias within Machine Learning Systems

Approved by:

_____, Thesis Committee Chair
Candy S. Smith, M.A., MMAS

_____, Member
David S. Pierson, Ph.D.

_____, Member
Brian L. Steed, Ph.D.

_____, Member
LTC Brandon C. Shelley, M.A.

Accepted this 10th day of June 2022 by:

_____, Assistant Dean of Academics for
Dale F. Spurlin, Ph.D. Degree Programs and Research

The opinions and conclusions expressed herein are those of the student author and do not necessarily represent the views of the U.S. Army Command and General Staff College or any other governmental agency. (References to this study should include the foregoing statement.)

ABSTRACT

AN EXPLORATION OF HOW THE MILITARY INTELLIGENCE COMMUNITY MAY MITIGATE UNINTENDED BIAS WITHIN MACHINE LEARNING SYSTEMS, by Bethany Bashor, 113 pages.

As the military modernizes with Artificial Intelligence/Machine Learning systems, the personnel in the military intelligence community will need to fully understand and integrate these systems into the Joint Intelligence Process (JIP). To guide the force in the development and integration of this technology, the Department of Defense (DOD) publicly released the "DOD Adopts Ethical Principles for Artificial Intelligence," which included "equitable" as a principle to mitigate "unintended bias." While there are multiple technical best practices to mitigate unintended bias, there are also nontechnical best practices that the military intelligence community could adopt from commercial industry. This qualitative, multiple case study and cross-case synthesis explores these commercial industry nontechnical best practices to mitigate unintended bias in the JIP.

ACKNOWLEDGMENTS

I want to express my sincere gratitude to my Command and General Staff College (CGSC) Master of Military Art and Science (MMAS) Thesis Committee. Without their expertise, guidance, and infinite patience, I would not have been able to complete this study. I am deeply appreciative for everything they taught me.

I also thank the CGSC Information Advantage Scholars Staff Group as well as my fellow students and the faculty within the CGSC MMAS thesis course. I am grateful that they took time out of their busy schedules to give me honest and constructive feedback. I valued their insight and advice as I navigated this process.

Lastly, I would like to acknowledge my husband. He repeatedly proofread my work without complaint and always cheered me throughout this entire endeavor. I am truly thankful for his support, patience, and kindness.

TABLE OF CONTENTS

Page

ACRONYMS

AI	Artificial Intelligence
DOD	Department of Defense
HK	Handcrafted Knowledge
IBM	International Business Machines
JAIC	Joint Artificial Intelligence Center
JIP	Joint Intelligence Process
ML	Machine Learning

ILLUSTRATIONS

Page

TABLES

Page

CHAPTER 1

INTRODUCTION

> By its nature intelligence is imperfect…
>
> —Chairman of the Joint Chiefs of Staff,
> Joint Publication 2-0, *Joint Intelligence*

Background

As part of an imperfect science and art, experts in the military intelligence community must consistently identify and mitigate bias in the pursuit of analytical "objectivity."[1] Identifying and mitigating bias can be particularly difficult if a machine, rather than a human, is executing the analytical task. The Department of Defense (DOD) is currently researching ways to utilize Artificial Intelligence (AI) systems to gain and maintain an "advantage" with information, and intelligence, in military operations.[2] In early 2020, the DOD published a "a series of ethical principles for the use of Artificial Intelligence" in its department.[3] As part of the DOD, the Joint Artificial Intelligence

[1] Chairman of the Joint Chiefs of Staff (CJCS), Joint Publication (JP) 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), II-8, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

[2] Department of Defense (DOD), *Summary of the 2018 Department of Defense Artificial Intelligence Strategy* (Washington, DC: Department of Defense, 2018), 11, https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

[3] Department of Defense (DOD), "DOD Adopts Ethical Principles for Artificial Intelligence," February 24, 2020, https://www.defense.gov/News/Releases/Release/Article/2091996/DOD-adopts-ethical-principles-for-artificial-intelligence/.

Center (JAIC) leads the effort for researching and integrating AI systems into the military, while simultaneously complying with these established ethics.[4]

While some may argue AI systems offer a way to overcome unintended bias, current AI systems are not sentient and rely upon humans for guidance. This means that bias will be designed into the AI system, and therefore, failing to reach objectivity in the subsequent intelligence.[5] The DOD recognized this concern and included "equitable" as one of its ethical standards.[6] By equitable, the "DOD should take deliberate steps to avoid unintended bias in the development and deployment of combat or non-combat AI systems that would inadvertently cause harm to persons."[7] Specifically, unintended bias refers to

[4] Joint Artificial Intelligence Center (JAIC), "Joint Artificial Intelligence Center," accessed November 21, 2021, https://www.ai.mil/.

[5] Greg Allen, "Understanding Artificial Intelligence Technology," Joint Artificial Intelligence Center, April 2020, 7-8, https://www.ai.mil/docs/Understanding%20AI%20Technology.pdf; John R. Allen and Darrell M. West, "How Artificial Intelligence is Transforming the World," The Brookings Institution, 2018, https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/.

[6] DOD, "DOD Adopts Ethical Principles for Artificial Intelligence."

[7] Defense Innovation Board, "AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense," (Supporting Document, U.S. Department of Defense, Washington, DC, October 2019), 31, https://media.defense.gov/2019/Oct/31/2002204459/-1/-1/0/DIB_AI_PRINCIPLES_SUPPORTING_DOCUMENT.PDF.

a flaw "that could undermine analytic validity and reliability, harm individuals…" as well as lead to "unintended outcomes." [8]

The DOD specified that "the term 'fairness' is often cited in the AI community," but the DOD does not seek fairness.[9] Instead, the "DOD aims to create the conditions to maintain an unfair advantage over any potential adversaries…."[10] As such, the DOD focuses on unintended biases that may be detrimental to friendly decision-making during military operations.[11]

Human AI technical designers writing unintended bias into their systems is a problem that spans both the psychological and technological fields. For the psychological aspect, technical designers must be aware of their own cognitive biases.[12] Human intelligence experts have continuously struggled with cognitive bias themselves. As such, the intelligence community developed various mitigation methods and techniques for its

[8] Office of the Director of National Intelligence (ODNI), "Artificial Intelligence Ethics Framework for the Intelligence Community," version 1 (Director of National Intelligence, Washington, DC, June 2020), 3, https://www.dni.gov/files/ODNI/documents/AI_Ethics_Framework_for_the_Intelligence_Community_10.pdf; Joint Artificial Intelligence Center (JAIC), "Department of Defense Joint Artificial Intelligence Center Responsible AI Champions Pilot," 2020, 4, https://www.ai.mil/docs/08_21_20_responsible_ai_champions_pilot.pdf.

[9] Defense Innovation Board, "AI Principles," 31.

[10] Ibid.

[11] Ibid.

[12] Osonde A. Osoba and Wesler William IV, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence* (Santa Monica, CA: RAND Corporation, 2017), i-35, https://www.rand.org/pubs/research_reports/RR1744.html.

experts to reference and utilize in their respective intelligence disciplines.[13] As for the technical aspect, technical designers must understand the specific AI biases, which can manifest separately from cognitive biases, as well as mitigation techniques. Designing AI systems with the military intelligence community could prove beneficial for technical designers in mitigating these unintended biases within an AI system.[14]

Between the two predominant types of AI systems, "Handcrafted Knowledge" (HK) and "Machine Learning" (ML), the DOD is currently focused on the implementation of ML systems into the force.[15] According to JAIC, the "key difference between a Handcrafted Knowledge System and a Machine Learning system is in where it receives its knowledge."[16] JAIC elaborates by stating, "rather than having their knowledge be provided by humans in the form of hand-programmed rules, Machine Learning systems generate their own rules."[17] An ML system is heavily reliant upon a well-written algorithm as well as diverse, accurate training data.[18] Both AI systems have advantages and disadvantages regarding their use and susceptibility to bias. The ML system is one of the AI systems that the military is currently trying to implement within

[13] Richards J. Heuer, Jr., *Psychology of Intelligence* (Langley,k VA: Central Intelligence Agency, Center for the Study of Intelligence, 1999), i-184, https://www.ialeia.org/docs/Psychology_of_Intelligence_Analysis.pdf.

[14] Osoba and Wesler, *An Intelligence in Our Image*, 17-24.

[15] Allen, "Understanding Artificial Intelligence Technology," 3.

[16] Ibid., 7.

[17] Ibid.

[18] Ibid., 7-8.

its strategic to tactical systems and processes.[19] This likely includes within the Joint Intelligence Process (JIP) which encompasses a cycle of "planning and direction; collection; processing and exploitation; analysis and production; dissemination and integration; and evaluation and feedback."[20]

Problem Statement

As the military continues to plan and implement ML systems within the force, the military intelligence community will need to understand indicators of and employ mitigation strategies against ML systems' unintended biases that may negatively influence the JIP.[21] Failure to mitigate these unintended biases may result in an ML system producing inaccurate intelligence, causing highly detrimental, or even fatal, consequences.

Purpose of the Study

The purpose of this study is to explore current commercial industry best practices that the military intelligence community can adopt to mitigate unintended bias in its ML systems. Several commercial industries have already embraced ML systems and developed several best practices for mitigating unintended biases in those systems.[22] This study analyzes and synthesizes some of these best practices to give the military intelligence community options when integrating its own ML systems. If the community

[19] Allen, "Understanding Artificial Intelligence Technology," 3.

[20] CJCS, JP 2-0, x.

[21] DOD, "DOD Adopts Ethical Principles for Artificial Intelligence."

[22] DOD, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 7-8.

implements these best practices, then this study may contribute to an increase in equitable intelligence support in a more expeditious manner.[23]

Research Questions

The primary research question for this study is as follows: how can the military intelligence community help mitigate unintended bias in an ML system operating in the JIP? The nested secondary research questions include:

1. How does unintended bias in ML systems influence situational understanding and decision-making?
2. What are the indicators of unintended bias in an ML system?
3. What are the best practices for mitigating unintended bias in an ML system?

Assumptions

This study recognizes multiple assumptions as true to continue research efforts. The first assumption is that the DOD will proceed with implementing ML systems into the military intelligence community.[24] The second assumption is that complete absence of unintended bias may not be attainable as humans are ultimately responsible for ML systems' design and guidance.[25] The third assumption is the DOD is still pursuing best practices on mitigation strategies against unintended biases within ML systems. The last

---

[23] Office of the Director of National Intelligence (ODNI), *The AIM Initiative: A Strategy for Augmenting Intelligence Using Machines* (Washington, DC: Director of National Intelligence, 2019), IV, https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf.

[24] DOD, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 1-17.

[25] Allen, "Understanding Artificial Intelligence Technology," 3.

assumption is that the DOD will continue to embrace a "human-centered" design method and is open to the concept of using "participatory design" best practices for adoption into its implementation strategies.[26]

Definition of Terms

There are several terms referenced throughout this study, and the definitions are below to ensure common understanding for the reader. This study provides definitions for "unintended bias" and "military intelligence community" which nest within the DOD's understanding of these concepts.

Artificial Intelligence is "the ability of machines to perform tasks that normally require human intelligence."[27]

Best Practice is "a procedure that has been shown by research and experience to produce optimal results and that is established or proposed as a standard suitable for widespread adoption."[28]

DOD "provides the military forces needed to deter war and ensure our nation's security."[29]

---

[26] DOD, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 6; Susan Gasson, "Human-Centered vs. User-Centered Approaches to Information System Design," *The Journal of Information Technology Theory and Application (JITTA)* 5, no. 2 (2003): 30.

[27] DOD, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 5.

[28] Merriam-Webster Dictionary Editors, "Best Practice," Merriam-Webster Dictionary, Merriam-Webster Incorporated, accessed January 25, 2021, https://www.merriam-webster.com/dictionary/best%20practice.

[29] Department of Defense (DOD), "U.S. Department of Defense," accessed October 12, 2021, https://www.defense.gov/.

Equitable refers to how the "the Department [DOD] will take deliberate steps to minimize unintended bias in AI capabilities."[30]

Human-Centered Design "is an approach to interactive systems development that aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, and usability knowledge and techniques."[31]

Intelligence Community "is composed of…18 organizations…" which consist of "two independent agencies…nine Department of Defense elements", to include the "intelligence elements of the five DOD services; the Army, Navy, Marine Corps, Air Force, and Space Force…" and "seven elements of other departments agencies…."[32]

Joint Intelligence Process is the cycle of "planning and direction; collection; processing and exploitation; analysis and production; dissemination and integration; and evaluation and feedback."[33]

Machine Learning is "the field of study interested in building computational systems that can improve their own performance of some task."[34]

---

[30] JAIC, "Department of Defense Joint Artificial Intelligence Center Responsible AI Champions Pilot," 4.

[31] International Organization for Standardization (ISO), "Ergonomics of Human-System Interaction—Part 210: Human-Centred Design for Interactive Systems," Online Browsing Platform, 2019, accessed May 5, 2022, https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:en.

[32] Office of the Director of National Intelligence (ODNI), "Members of the IC," accessed October 12, 2021, https://www.dni.gov/index.php/what-we-do/members-of-the-ic.

[33] CJCS, JP 2-0, x.

[34] ODNI, "The Aim Initiative," 15.

Military Intelligence Community refers to the intelligence personnel operating only within the DOD services.

Participatory Design is "an alternative to the traditional, technology-centered system development life-cycle that resulted from an emphasis on human-computer interaction…."[35]

Unintended Bias is a flaw "that could undermine analytic validity and reliability, harm individuals…" as well as lead to "unintended outcomes."[36]

Scope

While the DOD identified five AI ethical principles, this study only focuses on the equitability standard as it pertains to ML systems that may impact the military intelligence community.[37] It explores current commercial industry best practices to mitigate unintended biases in ML systems. As the DOD suggested that it will use a human-centered design approach, this study focuses only on best practices which fall under the overarching category of participatory design to complement the DOD's design approach.[38] In addition, this study only references best practices which may specifically benefit the military intelligence community.

---

[35] Gasson, "Human-Centered vs. User-Centered Approaches to Information System Design," 30.

[36] ODNI, "Artificial Intelligence Ethics Framework for the Intelligence Community," 3; JAIC, "Department of Defense Joint Artificial Intelligence Center Responsible AI Champions Pilot," 4.

[37] DOD, "DOD Adopts Ethical Principles for Artificial Intelligence."

[38] DOD, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 6.

Limitations

There are numerous limitations that apply to this study. The first limitation is time available, as the researcher only has approximately eight months to initiate and complete the research. Thus, the study does not explore all aspects of this topic. This leads to the second limitation which is the researcher has a basic understanding of ML. While the researcher has experience within the intelligence field, the researcher had to conduct extensive research to build a basic understanding of ML systems. If provided more time, the researcher may have been able to become more knowledgeable of the various technical nuances of ML systems; however, due to time constraints, knowledge was limited to what is feasible to learn within eight months. Additionally, due to time, the study does not focus on the technical aspects of ML systems, but instead focuses on participatory design best practices for the general military intelligence community. Lastly, the study references only English-speaking unclassified, publicly available information since access to proprietary or classified information was not available. This limitation restricts understanding of the innerworkings of a commercial industry's platform or application, which contains its ML systems, as this information is generally proprietary; therefore, the study approaches the referenced platform or application comprehensively as one whole ML system.

Delimitations

As there are limitations to this study, there are also delimitations. The study only refers to commercial industry best practices which may be employed by the military intelligence community to mitigate unintended biases. As such, the study focuses on participatory design best practices which the military intelligence community can

implement. Lastly, the study also only references current ML system best practices with the recognition that ML is a rapidly evolving field.

Significance of Study

As the DOD progresses with the implementation of ML systems, the military intelligence community must fully understand and integrate ML systems within its JIP while maintaining ethical standards.[39] Whether providing human or machine-derived intelligence, the military intelligence community has a responsibility to try to provide equitable intelligence to better inform senior leaders in their decision-making.[40] Failure to do so could have life and death consequences. This study explores current nontechnical commercial industry best practices which focus on participatory design mitigation strategies. The intent is to inform the military intelligence community on best practices it can implement throughout the JIP to mitigate unintended bias and provide better equitable intelligence to decision makers.

Summary

This chapter serves as the introduction for this study. It covers multiple aspects that span the background, problem statement, purpose, primary and secondary research questions, assumptions, terms and definitions, scope, limitations, delimitations, and significance. In summary, with the DOD's adoption of the AI ethical standard of equitability, the military intelligence community needs to understand how it can help

[39] DOD, "DOD Adopts Ethical Principles for Artificial Intelligence."

[40] Ibid.

mitigate unintended bias in ML systems used in the JIP.[41] To explore this topic, the primary and secondary questions for this study are as follows.

1. Primary research question: How can the military intelligence community help mitigate unintended bias in an ML system operating in the JIP??
2. Secondary research question: How does unintended bias in ML systems influence situational understanding and decision-making?
3. Secondary research question: What are the indicators of unintended bias in an ML system?
4. Secondary research question: What are the best practices for mitigating unintended bias in an ML system?

As this is an expansive topic, the scope of the study narrows to only how the military intelligence community can mitigate unintended bias in ML systems supporting the JIP. In addition, the study focuses on only commercial industry, participatory design best practices with current ML systems, not other AI systems. The following chapter reviews current literature and research pertinent to this study's problem statement as well as its primary and secondary research questions.

[41] DOD, "DOD Adopts Ethical Principles for Artificial Intelligence."

CHAPTER 2

LITERATURE REVIEW

Introduction

The concept of utilizing ML systems to augment a workforce is not new. Both the private and public sectors have pursued this technology since its creation. A major difference is how prevalent the concept of ML systems is today within U.S. national strategies, and the tenacity with which the government is currently pursuing this technology for its advantage.[42] This chapter will provide additional background on the military's Joint Intelligence Process (JIP), current AI efforts as well as an overview on ML systems. The chapter will then transition to current research and reporting regarding the study's secondary research questions, which focus on ML influence on situational awareness and decision-making, types and indicators of unintended bias, and best practices for mitigating bias in ML systems. After providing a comprehensive overview of current research relevant to this study, the chapter will conclude by identifying current gaps in literature and outlining the theoretical framework supporting this study's methodology.

[42] Allen, "Understanding Artificial Intelligence Technology," 3; DOD, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 6.

Background

Joint Intelligence Process

The JIP is a cyclical process that the military intelligence community uses to provide intelligence support to the military.[43] While each military service may have its own respective intelligence process, the JIP applies to the entire DOD; therefore, it is the basis for this study. There are six phases within the JIP, to include: "planning and direction; collection; processing and exploitation; analysis and production; dissemination and integration; and evaluation and feedback."[44] Each phase is distinct in its activities, but all are reliant upon one another to function properly. For the purposes of this study, the researcher uses the generic definition for each these phases. The first phase is planning and direction which is "best understood as the development of intelligence plans and the continuous management of their execution."[45] Collection is the second phase, and "includes those activities related to the acquisition of data required to satisfy the requirements…."[46] The next phase, processing and exploitation, is where "raw collected data is converted into forms that can be readily used by commanders, decision makers at all levels, intelligence analysts and other consumers."[47] The fourth phase is analysis and production, which is when "available processed information is integrated, evaluated,

[43] CJCS, JP 2-0, ix-x.

[44] Ibid., x.

[45] Ibid., I-5.

[46] Ibid., I-15.

[47] Ibid.

analyzed, and interpreted to create products…."[48] Once the fourth phase is complete, the "intelligence is delivered to and used by the consumer" during dissemination and integration.[49] The final phase is evaluation and feedback which happens "continuously throughout the intelligence process and as an assessment of the intelligence process as a whole."[50] This entire process is shown in Figure 1 below.

Figure 1. The Joint Intelligence Process

Source: Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

[48] CJCS, JP 2-0, I-16.

[49] Ibid., I-20.

[50] Ibid., I-21.

Figure 2 below demonstrates how three of these phases – "collection," "processing and exploitation," and "analysis and production" – filter and process data to become intelligence.[51] Essentially, these phases each serve as a gateway to the subsequent phases and help identify which data and information is pertinent to the military. If humans or ML systems execute these phases incorrectly, then this can have negative consequences in the subsequent steps and change a final analytical assessment. It may also substantially slow the cycle due to confusion from incomplete information as well as troubleshooting to identify where the error occurred.[52]

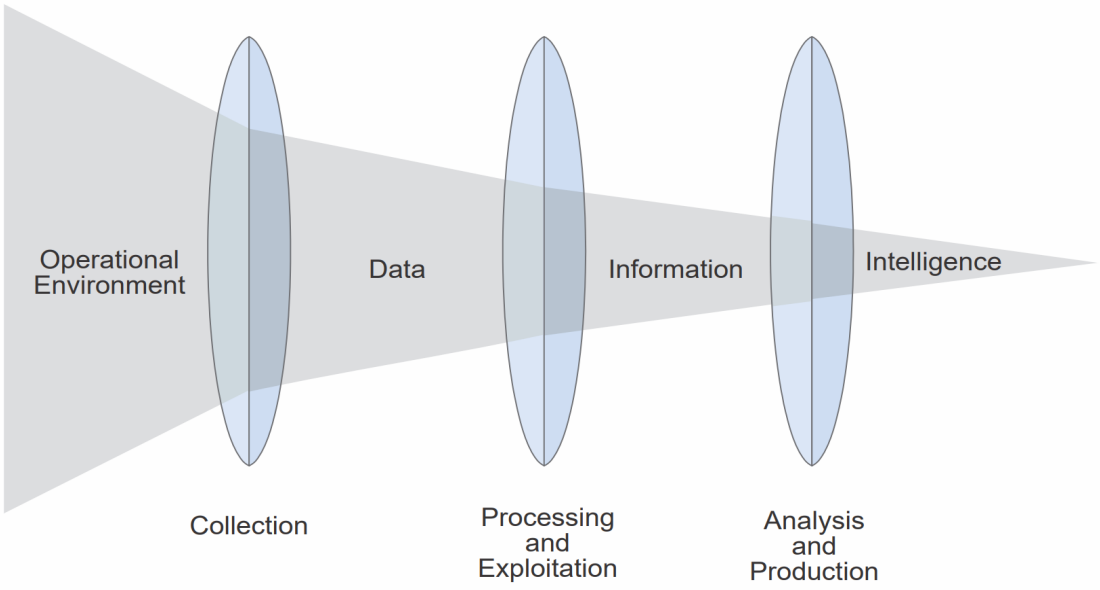

Figure 2. How Intelligence is Filtered from Data

Source: Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-2, figure I-1, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

[51] CJCS, JP 2-0, I-2, fig. I-1.

[52] Ibid., I-1 to I-23.

Military AI Efforts

While the military has worked with AI for several decades, it began focusing modernization efforts to integrate ML systems into the force upon the publication of the 2017 *National Security Strategy of the United States of America*.[53] In 2018, the Secretary of Defense published the *Summary of the National Defense Strategy*, which reiterated the importance of developing ML system capabilities for the military's use.[54] In the same year, the DOD also published its *Summary of the 2018 Department of Defense Artificial Intelligence Strategy* where it identified the JAIC as the leader for this effort.[55] In addition, the DOD requested that the Defense Innovation Board (DIB) establish an AI ethical framework, which the DOD adopted in 2020.[56] One of ethical standards within this framework was "equitable."[57] The DOD uses the term equitable to describe the standard to identify and mitigate unintended bias with ML systems; however, equitable is not equivalent to fairness as the DOD seeks "an unfair advantage over any potential adversaries…."[58] One of the latest publications, the *DOD AI Education Strategy*, outlines

---

[53] U.S. President, *National Security Strategy of the United States of America* (Washington, DC: The White House, 2017), i-55, https://history.defense.gov/Portals/70/Documents/nss/NSS2017.pdf?ver=CnFwURrw09pJ0q5EogFpwg%3d%3d.

[54] Secretary of Defense, *Summary of the 2018 National Defense Strategy* (Washington, DC: Department of Defense, 2018), 7, https://DOD.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf.

[55] DOD, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 9.

[56] DOD, "DOD Adopts Ethical Principles for Artificial Intelligence."

[57] Defense Innovation Board, "AI Principles," 31.

[58] Ibid.

how JAIC envisions the implementation of AI education throughout the force. Notably, it separates personnel into "archetypes" which include "lead," "drive," "create," "embed," "facilitate," and "employ" AI.[59] Figure 3 below provides an overview of each of these archetypes and how they fit within the overall implementation of AI/ML systems.[60]

Archetype	Description	Concentration	Role Explanation
Lead AI	Decides policy and doctrine, including how AI tools can or will be used; builds AI vision and plan	Policy	• Creates overarching guidance on DOD AI use
		Command	• Ensures AI policy carried out by personnel they lead
		Agency/Function Lead	• Ensures AI policy carried out in non-combat agencies
Drive AI	Ensures appropriate AI tools and capabilities are developed and delivered across DOD	Acquisitions Manager	• Supports technology/capabilities through total life cycle
		Capability Manager	• Evaluates and develops force structure resources and reqs
		Technical Manager	• Defines the tech strategy across a project portfolio
		Product Manager	• Ensures the creation of AI-enabled tools, from start to finish
Create AI	Creates AI tools to meet current and future needs	AI Researcher	• Pushes DoD AI capability by preparing for future use cases
		AI/ML Engineer	• Builds, tests, codes, integrates, and delivers AI tools
		Testing & Evaluation Engineer	• Evaluates system capabilities, limitations, operational risks
		Data Scientist	• Applies AI tools to perform analytics and create solutions
		Deployment Engineer	• Manages integration, deployment, and operation of AI systems
Embed AI	Embedded with Employ AI, establishes AI systems and provides end-user support at tactical edge	Technician	• Deploys, maintains, adapts, and collects data for AI/ML systems at the tactical edge
Facilitate AI	Represents users to ensure appropriate AI tools are developed and delivered to address use cases	Product Owner	• Provides voice of customer; turns product vision into backlog
		UI/UX	• Designs AI tool interface for usability and accessibility
		Other Technical Experts	• Delivers discrete elements of system not specific to AI
Employ AI	End-users of AI tools, provide feedback on and requirements for AI tools	Operations	• Prepares for and delivers operational requirements
		Intelligence	• Gathers and analyzes info to support decision-making
		Logistics & Maintenance	• Enables troop / gear movement, maintain equipment
		Health	• Maintains health and wellbeing of the Warfighter
		Support	• Supports the Warfighter in non-combat requirements

Figure 3. The DOD AI Education Strategy Archetypes

Source: Joint Artificial Intelligence Center DoD Chief Information Officer, "2020 Department of Defense Artificial Intelligence Education Strategy," (Joint Artificial Intelligence Center, Washington, DC, September 2020), 7, https://www.ai.mil/docs/2020_DoD_AI_Training_and_Education_Strategy_and_Infographic_10_27_20.pdf.

[59] Joint Artificial Intelligence Center (JAIC) DoD Chief Information Officer, "2020 Department of Defense Artificial Intelligence Education Strategy," (Joint Artificial Intelligence Center, Washington, DC, September 2020), 7, https://www.ai.mil/docs/2020_DoD_AI_Training_and_Education_Strategy_and_Infographic_10_27_20.pdf.

[60] Ibid.

ML System Overview

Designers may introduce unintended bias into an ML system in at least two ways: through its algorithm and its data. An ML system is algorithmically based and trained on certain data sets, "by running a human-generated algorithm on the training dataset, the Machine Learning system generates the rules…" to execute its mission.[61] This training data is varied, but generally focuses on "supervised," "unsupervised," "semi-supervised," or "reinforcement."[62] The difference between supervised and unsupervised training is whether a designer "labels" all the training data for the ML system.[63] "Supervised Learning uses example data that has been labeled by…[designers]. Supervised Learning has incredible performance, but getting sufficient labeled data can be difficult, time-consuming, and expensive."[64] The opposite of supervised training is unsupervised training, which uses non-labeled data. This type of training is not as effective as supervised training; however, it can be extremely helpful when it is unrealistic to gather large amounts of labeled data or manually label all of it. The next type of training, semi-supervised, uses a mix of "labeled and unlabeled data."[65] As with unsupervised training, this is not as effective as supervised training, but can be beneficial when one needs to process a mix of data quickly and has limited human resources.

---

[61] Allen, "Understanding Artificial Intelligence Technology," 7.

[62] Ibid., 4.

[63] Ibid.

[64] Ibid.

[65] Ibid.

The last type of training is reinforcement. This is where an ML system "gather[s] their own data and improve[s] based on their trial and error interaction with the environment."[66] While this is perhaps what most people think of when they think of AI, this type of training is still undergoing extensive research.[67] Based on this information, it is easier to understand how designers may introduce unintended bias into an ML system. For example, if a designer mislabels data, or labels data according to his or her own unintended biases, then this could skew how the ML system learns to interpret the data in the future. In addition, if a designer writes the ML system algorithm to look for only certain items, then the ML system could potentially overlook other key items that are crucial for situational understanding.[68]

ML Influence on Situational Understanding and Decision-Making

ML systems could have the potential to impact intelligence assessments, and consequently, situational understanding and decision-making. Current research indicates that whether an ML system is influential generally depends on whether humans trust it; however, when trusted, it can influence situational awareness, and subsequently, decision making.[69] One of the most effective ways to build this trust is through education and

[66] Allen, "Understanding Artificial Intelligence Technology," 4.

[67] Ibid.

[68] Ibid., 1-20.

[69] IBM, "Building Trust in AI," accessed February 11, 2022, https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/building-trust-in-ai.html.

clarifying the capabilities of ML systems.[70] As such, the DOD recognized this and already published its *DOD AI Education Strategy* with the intent to increase AI literacy in the force.[71]

ML systems are highly prevalent throughout industry, and various studies have shown how ML systems can enhance human situational awareness and decision-making within multiple business sectors.[72] Within finance, it helps businesses complete a range of activities, such as determining candidacy for loans, detecting fraud, and improving cybersecurity.[73] This differs widely from the medical field's exploration of using ML systems for "decision support (disease diagnosis and screening), processing patient medical data (e.g., detecting abnormalities in an X-ray or fundus image), or optimizing processes and services."[74] Alternatively in supply chain management, businesses are exploring ways to use ML systems to build "intelligent workflows to improve visibility throughout..[their] supply chain[s]."[75] In these few examples, ML systems impacted how

---

[70] Ibid.

[71] JAIC, "2020 Department of Defense Artificial Intelligence Education Strategy," i-49.

[72] Allen and West, "How Artificial Intelligence is Transforming the World."

[73] Lael Brainard, "What Are We Learning about Artificial Intelligence in Financial Services?" (Speech, Fintech and the New Financial Landscape, Philadelphia, Pennsylvania, November 13, 2018), https://www.federalreserve.gov/newsevents/speech/brainard20181113a.htm.

[74] Richard Fletcher, Audace Nakeshimana, and Olusubomi Olubeko. "Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health," *Frontiers in Artificial Intelligence* 3 (April 2021), https://doi.org/10.3389/frai.2020.561802.

[75] IBM, "Build Smarter Supply Chains with AI and Blockchain," accessed May 1, 2022, https://www.ibm.com/supply-chain.

those businesses processed information and made subsequent decisions. While issues have surfaced with the use of these systems in these fields, especially regarding unintended bias, businesses continue to pursue the use of ML systems as they foresee opportunity. ML systems offer the opportunity for these fields to increase "processing speeds," increase output, and focus the workforce on human specific needs.[76] As such, it can also enhance situational awareness and increase effective and efficient decision making, which for businesses is critical to remaining competitive with peers.[77]

Types and Indicators of Unintended Bias in ML Systems

As with any tool, there will always be advantages and disadvantages in its use. While many see these ML systems with an "aura of objectivity and infallibility" to process information, studies indicate that these systems suffer heavily from unintended bias.[78] Some studies have analyzed a variety of ways to identify unintended bias within an ML system. Generally, these studies indicate that this bias is rooted in the algorithm, its training, or the evaluation of the ML system.[79] The military intelligence community needs to be able to identify the indicators and types of unintended biases within ML systems which are operational in the JIP. With the capability to identify unintended bias

[76] Allen and West, "How Artificial Intelligence is Transforming the World."

[77] Ibid.

[78] Osoba and Wesler, *An Intelligence in Our Image*, iii.

[79] Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 45-64, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134; Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan, "A Survey on Bias and Fairness in Machine Learning," (Arxiv, Cornell University, 2022), 1-34, https://arxiv.org/pdf/1908.09635.pdf.

in ML systems, the community can conduct oversight and improve the ML systems as well as provide the most accurate intelligence possible to decision makers. Without this capability, bias may continue to permeate the ML system and negatively influence decision making and situational understanding in the JIP. While not an exhaustive list, below are a few key types and indicators of unintended biases that may manifest in ML systems.

There are a variety of unintended biases with which the military intelligence community must contend when using ML systems. While there is not a formalized list of unintended biases, authors Ajay Chander and Ramya Srinivasan identified several common ones in their study "Biases in AI Systems," which include "sampling bias…measurement bias…label bias…negative set bias…framing effect bias…sample selection bias…confounding bias…design bias…sample treatment bias…human evaluation bias…[and] test dataset biases."[80] These biases are found throughout the various stages of the "AI/ML pipeline" and contain several subcategories of unintended biases within each of them.[81]

During the initial stage of creating a ML system, one must first compile data in which the unintended biases of sampling, measurement, label, and negative set are most prevalent.[82] Sampling bias "arises in a dataset that is created by selecting particular types of instances more than others (and thereby rendering the dataset underrepresentative of

[80] Chander and Srinivasan, "Biases in AI Systems," 56.

[81] Ibid.

[82] Ibid., 48-50.

the real world)."[83] This may manifest as an indicator when "trends estimated for one population may not generalize to data collected from a new population."[84] The next bias is measurement bias which "is introduced by errors in human measurement, or because of certain intrinsic habits of people in capturing data."[85] An indicator of this bias might be a trend of the ML system to only display certain types of data in a particular way due to limited data measurements.[86] As for label bias, it is "associated with inconsistencies in the labeling process."[87] This is relatively simple to identify as it usually manifests with data populating with incorrect or inconsistent labels.[88] The last bias specific to data collection is negative set bias which is described "as being introduced in the dataset as a consequence of not having enough samples representative of 'the rest of the world'."[89] An indicator of this type of bias is when a ML system classifier is unable to identify a piece of data by "what it is not (negative instances)."[90]

---

[83] Chander and Srinivasan, "Biases in AI Systems," 48.

[84] Mehrabi et al., "A Survey on Bias and Fairness in Machine Learning," 6.

[85] Chander and Srinivasan, "Biases in AI Systems," 49.

[86] Ibid.

[87] Ibid.

[88] Ibid.

[89] Ibid., 50.

[90] Antonio Torralba and Alexei Efros, "Unbiased Look at Dataset Bias," in *IEEE Conference Proceedings on Computer Vision and Pattern Recognition,* IEEE, June 2011, 1521-1528; https://ieeexplore.ieee.org/document/5995347, quoted in Chander and Srinivasan, "Biases in AI Systems," 50.

The next set of unintended biases focuses on "problem formulation" and "data analysis."[91] The bias specific to problem formulation is framing effects bias which is "based on how the problem is formulated and how information is presented, the results obtained can be different and perhaps biased."[92] An indicator of this bias is when a ML system provides inconsistent query results between groups of similar data.[93]

The following biases fall within the data analysis category. Sample selection bias "is introduced by the selection of individuals, groups, or data for analysis in such a way that the samples are not representative of the population intended to be analyzed."[94] Confounding bias is where an "algorithm learns the wrong relations by not taking into account all the information in the data or if it misses the relevant relations between features and target outputs."[95] One indicator of this type of bias includes omissions of data.[96] Another indicator is the use of a "proxy variable" where the designers purposely omit "sensitive variables," but the ML system utilizes another variable as a substitute.[97] Next is design bias, which includes its sub-types of "algorithm," "ranking," and "presentation" biases.[98] Algorithm bias is bias "induced or added by the algorithm."[99]

---

[91] Chander and Srinivasan, "Biases in AI Systems," 51-52.

[92] Chander and Srinivasan, "Biases in AI Systems," 51.

[93] Ibid.

[94] Ibid., 52.

[95] Ibid.

[96] Ibid., 53.

[97] Ibid.

[98] Ibid., 56.

This type of bias can be difficult to identify and requires designers to review the "algorithmic design choices" for correction.[100] Ranking bias refers to when a ML system ranks certain items which inherently gives "privilege" to certain items over others.[101] Presentation bias is similar to ranking bias, and specifically refers to how one can only "receive user feedback only on items that have been presented to the user."[102]

The last category of unintended biases focuses on "evaluation/validation."[103] One of the most familiar types of bias is the human evaluation biases.[104] This occurs when humans "are employed in validating the performance of an AI model. Phenomena such as confirmation bias, peak end effect, and prior beliefs (e.g., culture) can create biases in evaluation."[105] This type of bias also exists outside of ML systems and indicators range widely; however, an indicator includes the "use of inappropriate and disproportionate benchmarks for evaluation of applications."[106] The last unintended bias is the sample treatment bias which is "the bias introduced in the process of selectively subjecting some

---

[99] Ibid., 53.

[100] Mehrabi et al., "A Survey on Bias and Fairness in Machine Learning," 7.

[101] Chander and Srinivasan, "Biases in AI Systems," 54.

[102] Ibid.

[103] Ibid.

[104] Ibid.

[105] Ibid.

[106] Mehrabi et al., "A Survey on Bias and Fairness in Machine Learning," 8.

sets of people to a type of treatment."[107] This includes an indicator such as revealing data to one group of people and not showing it to others.[108]

Best Practices for Mitigating Unintended Bias in an ML System

Just as there are several different indicators of unintended bias within an ML system, there are a multitude of various best practices mitigation strategies. These strategies generally fall within two categories, "technical and nontechnical" approaches.[109] The type and severity of the unintended bias within the ML system will generally dictate which method to employ.[110]

For technical mitigation strategies, the preponderance of work falls upon the designers, as they have access to the algorithm and data. Upon identification or notification of an indicator of unintended bias, the designer should either review to recode the algorithm or adjust the training data to correct the ML system.[111] There are also "toolkits" which are designed to continuously "examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle."[112] For example, an International Business Machines (IBM) toolkit is available in open source and uses the public to help identify and mitigate unintended bias within

---

[107] Chander and Srinivasan, "Biases in AI Systems," 55.

[108] Ibid., 54-55.

[109] Osoba and Wesler, *An Intelligence in Our Image*, 21.

[110] Ibid.

[111] Ibid., 21-22.

[112] IBM, "AI Fairness 360," accessed November 16, 2021, https://aif360.mybluemix.net/.

ML systems.[113] In addition, as humans write an ML system's algorithm and design its dataset, regardless if supervised or unsupervised, "diversity in the ranks of algorithm developers [designers] could help improve sensitivity to potential disparate impact problems."[114]

While technical means are likely preferred, as it implies that the designers can directly fix the ML system's algorithm or data, nontechnical mitigation strategies can also be very beneficial. As such, businesses are starting to explore participatory design best practices to mitigate bias. "Participatory artificial intelligence or participatory machine learning in their broadest sense refer to the involvement of a wider range of stakeholders than just technology developers in the creation of an AI system, model, tool or application."[115] For the purposes of this study, stakeholder refers to "a person such as an employee, customer, or citizen who is involved with an organization, society, etc. and therefore has responsibilities toward it and an interest in its success."[116] One can further subcategorize participatory design into four main best practices which are "co-creation," "collaboration," "contribution," and "consultation."[117] Stakeholder participation is the

---

[113] Ibid.

[114] Osoba and Wesler, *An Intelligence in Our Image*, 24.

[115] Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, London, UK, 2021), 6, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf.

[116] Cambridge Dictionary Editors, "Stakeholder," Cambridge Dictionary, Cambridge University Press, accessed April 18, 2022, https://dictionary.cambridge.org/us/dictionary/english/stakeholder.

[117] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 11.

highest in co-creation, and it steadily decreases in the remaining three. Co-creation has "external stakeholders help to initiate and oversee the project as well as collaborating throughout the AI pipeline."[118] For example, a team that consists of both technical developers and stakeholders to design a ML system would be co-creation.[119] Next is collaboration which is where external stakeholders' "participation has multiple touchpoints along the AI pipeline and includes activities where the technical team interacts with participants."[120]

An example of this best practice would be having key stakeholders who would have regularly scheduled interactions with the technical developers but do not participate in the daily design process with the technical developers.[121] The third participatory design best practice is contribution in which "participation is time-limited and involves stakeholders completing tasks that are necessary to AI development. Participants do not interact with each other or the technical team."[122] This best practice can appear in the form of "crowdsourcing" where stakeholders assist with a specific part of the ML system design process.[123] An example of this best practice is when stakeholders, or even a

---

[118] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 11.

[119] Ibid., 13.

[120] Ibid., 11.

[121] Ibid., 13.

[122] Ibid., 11.

[123] Jennifer Vaughan, "Making Better Use of the Crowd: How Crowdsourcing can Advance Machine Learning Research," *Journal of Machine Learning Research*, 18 (2018), 1, https://www.jmlr.org/papers/volume18/17-234/17-234.pdf.

broader swath of users, assist with "labelling" or categorizing data for an ML system.[124] The last best practice is consultation where "input occurs outside of the core AI development process and it is not guaranteed that it will impact the design of the AI."[125] For example, "focus groups and interviews" are popular with "technical teams or organisations" to find any issues from users.[126] "The goals of consultations are typically twofold; some initiatives seek to understand public attitudes for different applications of AI or regulatory policy, while others aim to increase data literacy and raise awareness of the impacts of AI systems."[127]

Regardless of technical or nontechnical measures, AI literacy is critical for all stakeholders to ensure that ML systems operate properly. The DOD already identified this with its *AI Education Strategy*.[128] As part of its outlined curriculum, the DOD acknowledged that not only is it important to educate users on how to properly use the ML system, but also to understand the system's capabilities and limitations.[129] In addition, educated users or stakeholders could have great potential to properly inform technical designers of any flaws in the system.[130]

---

[124] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 12.

[125] Ibid., 11.

[126] Ibid., 12.

[127] Ibid., 11.

[128] JAIC, "2020 Department of Defense Artificial Intelligence Education Strategy," i-49.

[129] Ibid., ii.

[130] Osoba and Wesler, *An Intelligence in Our Image*, 23-24.

Gaps in Literature

There has been substantial research on identifying and mitigating unintended bias in ML systems within industry and the public sphere; however, there is much to explore regarding how best to do this within the military intelligence community. As this capability is relatively new, and expanding rapidly, the military intelligence community needs to understand how it could apply commercial industry best practices to its own systems and processes to try to mitigate unintended bias within the JIP. While most of the military intelligence community might not have the technical understanding of ML systems, additional research may be able to highlight approaches, specifically nontechnical methods, it could take to help mitigate unintended bias in ML systems in the JIP.

Theoretical Framework

"Computational Learning Theory," also known as "Machine Learning Theory," is the predominate underlying theory to this study.[131] This theory seeks "to understand the basic algorithmic principles involved in getting computers to learn from data and to improve performance with feedback…."[132] Understanding this part of the theory is critical as this study seeks to improve ML systems by exploring how the military intelligence community can mitigate unintended biases by implementing participatory design best practices. As such, this study takes the perspective of this specific community

[131] Avrim Blum, "Machine Learning Theory" (Essay, Carnegie Melon University, School of Computer Science, 2007), 1, http://www.cs.cmu.edu/afs/cs/user/avrim/www/Talks/mlt.pdf.

[132] Ibid., 2.

in trying to mitigate unintended bias in its ML systems. Subsequent research focuses on participatory design best practices from the commercial industry that are relatable to the intelligence field. In its essence, this study's theoretical framework assumes the perspective of the general military intelligence community, not the technical designer, on how best to use participatory design best practices to improve ML systems and ensure it can complete its mission in accordance with ethical standards as established by the DOD.[133]

Summary

This chapter provides a broad, comprehensive overview of the current literature on how ML systems currently influence situational understanding and decision-making as well as some of the indicators of and current best practices to mitigate unintended bias within these systems. This literature review provides an understanding of how best to shape this study's research methodology in that a considerable amount of the related research is qualitative in nature, with the majority focusing on case studies. This is understandable, as ML systems can be employed in a variety of ways throughout multiple fields, which also influences how, and which unintended biases may appear. As such, not all the best practices may apply in every situation. With this understanding of current research, the researcher conducts a qualitative case study with the knowledge that not all best practices discussed may apply to every ML system used by the military intelligence community. The next chapter will discuss the overall research methodology for this study.

[133] DOD, "DOD Adopts Ethical Principles for Artificial Intelligence."

CHAPTER 3

RESEARCH METHODOLOGY

Introduction

This study's qualitative methodology comprises a two-part process, a multiple case study analysis followed by a cross-case synthesis. The primary research question focuses on how the military intelligence community can help mitigate an ML system's unintended bias in the JIP. The nested secondary research questions include: 1) how does unintended bias within ML systems influence situational understanding and decision-making? 2) what are the indicators of unintended bias in an ML system? 3) what are the best practices for mitigating unintended bias in an ML system? The literature review reveals a qualitative case study is the most appropriate methodology for answering these primary and secondary research questions. As not every best practice mitigation strategy is present in every case study, the researcher collects, analyzes, and cross references multiple case studies from commercial industry for individual analysis and cross-case synthesis.[134] This chapter covers the following topics in order: method, data collection, data analysis, and the summary.

[134] Robert Yin, *Case Study Research and Applications* (Thousand Oaks, CA: SAGE Publications, Inc., 2018), 3-23.

Method

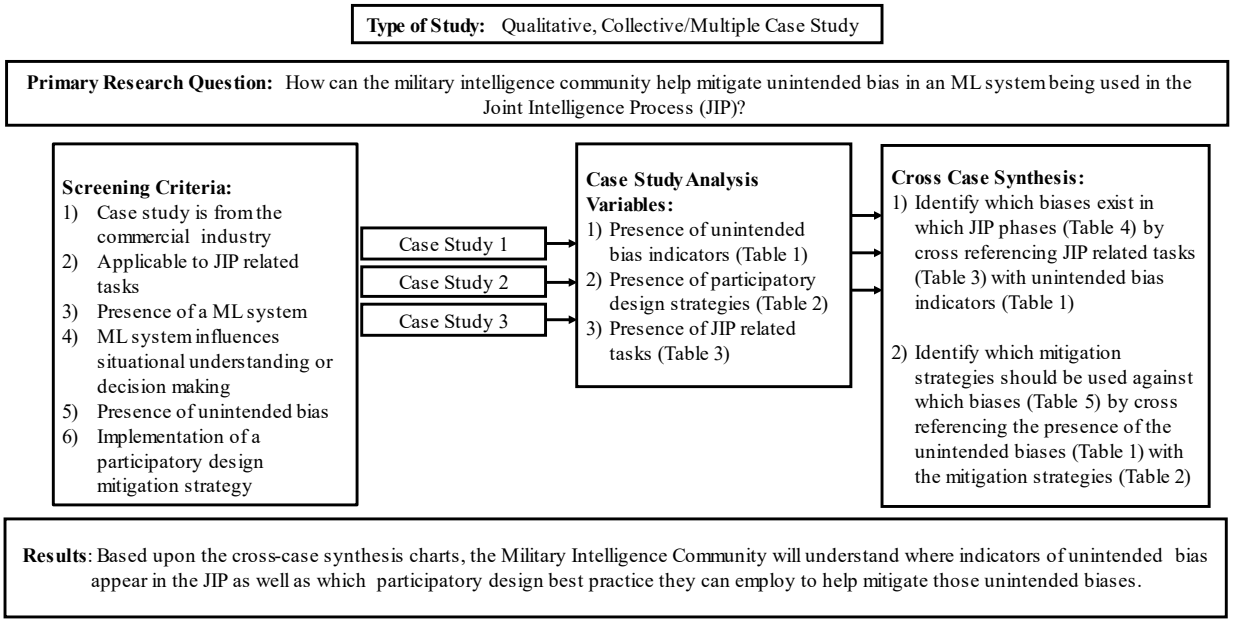

Figure 4. Methodology Workflow

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134; Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, London, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf; Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

According to Robert Yin in his book, *Case Study Research and Applications,* a qualitative case study is appropriate when it fits three predominant categories: 1) when the study is trying to answer a "how" or "why" primary research question, 2) does not

have "control over behavior events," and 3) "focuses on contemporary events."[135] Primary research questions focused on the *how* and *why* "are more explanatory and likely to lead to the use of a case study, history, or experiment as the preferred research method."[136] This study's primary research question focuses on the *how,* which meets Yin's first criteria for a case study. Secondly, Yin further explains case studies are preferable "when the relevant behaviors still cannot be manipulated and when the desire is to study some contemporary event or set of events ('contemporary' meaning a fluid rendition of the recent past and the present, not just the present)."[137] This differentiates case studies from an experiment where researchers have control.[138]

This study analyzes how the commercial industry has implemented participatory design best practices to mitigate unintended bias in its ML systems, and therefore, the researcher has no control over these businesses' decisions. This satisfies Yin's second criteria for a case study. Lastly, while a history study also does not have control, it differs from a case study as it does not focus on contemporary events.[139] As such, Yin's criteria eliminates both a history and an experiment as possible methods for this study, leaving the case study as the preferable method.

---

[135] Yin, *Case Study Research and Applications,* 9.

[136] Ibid., 10.

[137] Ibid., 12.

[138] Ibid., 9.

[139] Ibid.

Within the case study methodology, there are also subcategories, "single" versus "multiple" cases.[140] The researcher chose to conduct a multiple case study due to the variation in types and indicators of unintended bias, subcategories of participatory design best practices, and applicability to JIP related tasks. Additionally, Yin argues that "multiple-case designs may be preferred over single-case designs" as they strengthen a study.[141] As such, a multiple case study would be the most appropriate method. To identify appropriate case studies, the researcher establishes a set of screening criteria. This requires the case studies to meet the following criteria: they 1) originate from commercial industry, 2) are applicable to JIP related tasks, 3) include the presence of a ML system, 4) involve a ML system used for situational awareness or decision making, 5) include the presence of an unintended bias, and 6) include the implementation of a participatory design best practice mitigation strategy. The screening criteria is reflected in the red box in Figure 5 below.

[140] Yin, *Case Study Research and Applications,* 61.

[141] Ibid.

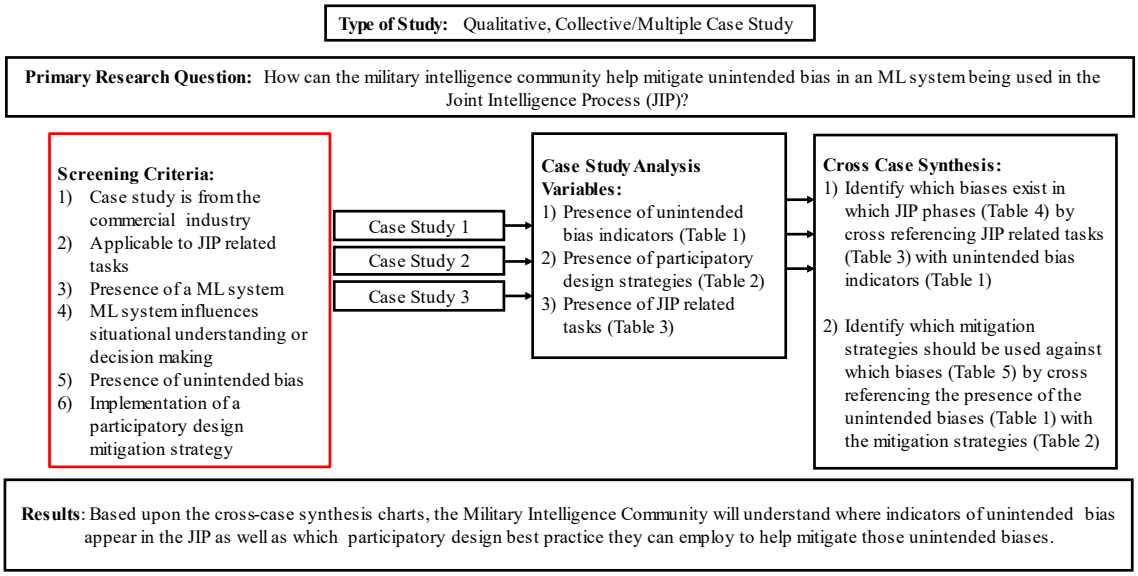

Figure 5. Methodology Workflow: Screening Criteria

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134; Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, Londong, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf; Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

After identifying three case studies, the researcher analyzes each case study for the presence of three variables. These variables are unintended bias indicators, participatory design best practice mitigation strategies, and JIP related tasks. This step is reflected in the red box in Figure 6 below. Upon completion of the individual case study analysis, the researcher cross references these three variables in a cross-case synthesis. This step is reflected in the red box in Figure 7 below. The intention is to help military intelligence community identify indicators of unintended bias throughout the JIP and

which participatory design best practice mitigation strategies it can employ against certain unintended biases.

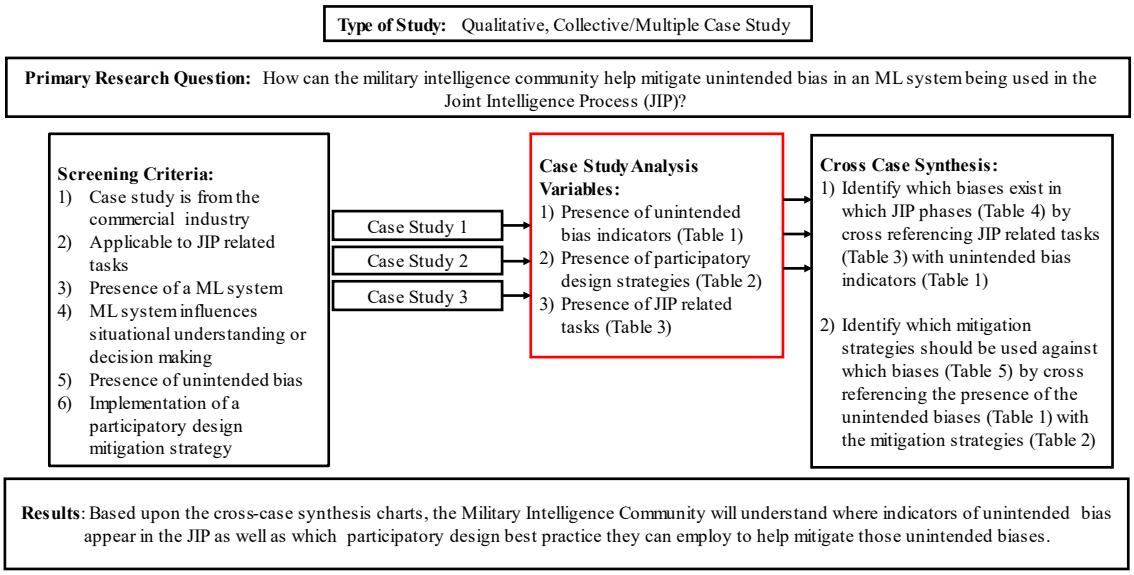

Figure 6. Methodology Workflow: Case Study Analysis Variables

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134; Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, Londong, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf; Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

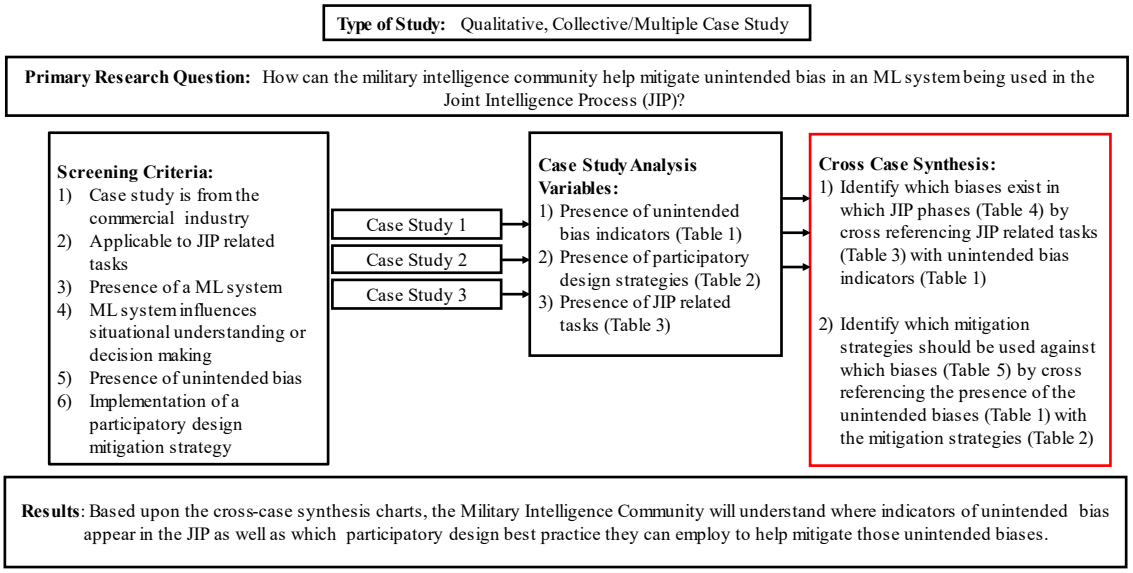

Figure 7. Methodology Workflow: Cross-Case Synthesis

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134; Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, Londong, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf; Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

Data Collection

For this study, the researcher collects three case studies from publicly available documentation where commercial industry identified unintended bias and implemented a participatory design best practice mitigation strategy. Due to the limitation of relying upon publicly available information, most of the collected documentation originates from academia or news publications as well as public formal, and informal (e.g., business

blogs), announcements from the respective businesses. Lastly, as per the study's design, the researcher does not engage with human subjects for this study.

Data Analysis

For each individual case study, the researcher analyzes the materiel for the presence of three variables from the perspective of the military intelligence community. These variables are as follows: 1) presence of unintended bias indicators, 2) presence of participatory design best practice mitigation strategies, and 3) presence of JIP related tasks. Tables one, two, and three are the templates for displaying the results of the case study analysis. For every positive indication of presence, the study provides citations from the analyzed documentation. The study specifically refers to the key terms and definitions of unintended bias, types of participatory design best practice mitigation strategies, and JIP phases from the literature review to determine if there is a presence within the case studies.

Regarding the JIP phases, the researcher acknowledges the terminology in these definitions are military-specific; however, the overall concepts are applicable to commercial industry. The researcher determines if the unintentionally biased ML system is conducting JIP related tasks by using generic definitions, derived from the JIP phase definitions in the literature review, provided in this section. Of note, as per the study's limitations, the researcher looks at the ML system's activities wholistically due to lack of access to proprietary information to discern exactly how the ML system functions. For a ML system's role in the first phase, planning and integration, the study defines this role as the ML system assisting with "the development of…plans and the continuous

management of their execution…."[142] Next is the second phase, collection, where a ML system assists with the "the acquisition of data required to satisfy the requirements…" of the consumer.[143] When referring to "acquisition," this means the "process of getting something" whether by taking or receiving it.[144] In this phase, the ML system may serve as the "collection manager" and/or the "collection asset."[145] A collection manager is the entity responsible for overseeing the "the process of converting…requirements into collection requirements, establishing priorities, tasking or coordinating with appropriate collection sources or agencies, monitoring results, and retasking, as required."[146] The collection asset refers to the actual device or entity that a collection manager chooses as "best suited to collect the information needed…."[147] In the third phase, processing and exploitation, the ML system helps facilitate that the "raw collected data is converted into forms that can be readily used by…consumers."[148] As for the fourth phase, the study defines this as where ML system may be able to assist with comprehensive analysis and production, which is when "processed information is integrated, evaluated, analyzed, and

[142] CJCS, JP 2-0, I-5.

[143] Ibid., I-15.

[144] Cambridge Dictionary Editors, "Acquisition," Cambridge Dictionary, Cambridge University Press, accessed April 15, 2022, https://dictionary.cambridge.org/us/dictionary/english/acquisition.

[145] CJCS, JP 2-0, I-13 to I-14.

[146] Ibid., I-13.

[147] Ibid., I-14.

[148] Ibid., I-15.

interpreted to create products that will satisfy…" consumer requirements or questions.[149] In the fifth phase, dissemination and integration, the ML systems helps ensure the products are "delivered to and used by the consumer."[150] The sixth and final phase, evaluation and feedback, is where the ML system assists with providing an "assessment" of the execution of the first five phases and the impact.[151]

Table 1. Case Study Variable 1 Template

Bias Group	Bias Type	Case Study 1:	Case Study 2:	Case Study 3:
Data Creation	Sampling	Y/N	Y/N	Y/N
	Measurement	Y/N	Y/N	Y/N
	Label	Y/N	Y/N	Y/N
	Negative Set	Y/N	Y/N	Y/N
Problem Formulation	Framing Effect	Y/N	Y/N	Y/N
Data Analysis	Sample Selection	Y/N	Y/N	Y/N
	Confounding	Y/N	Y/N	Y/N
	Design	Y/N	Y/N	Y/N
Evaluation & Validation	Sample Treatment	Y/N	Y/N	Y/N
	Human Evaluation	Y/N	Y/N	Y/N
	Validation and Test Dataset	Y/N	Y/N	Y/N

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134.

[149] CJCS, JP 2-0, I-16.

[150] Ibid., I-20.

[151] Ibid., I-21.

Table 2. Case Study Variable 2 Template

Participatory Design	Case Study 1	Case Study 2	Case Study 3
Co-Creation	Y/N	Y/N	Y/N
Collaboration	Y/N	Y/N	Y/N
Contribution	Y/N	Y/N	Y/N
Consultation	Y/N	Y/N	Y/N

Source: Created by author using information from Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, London, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf.

Table 3. Case Study Variable 3 Template

JIP Phases	Case Study 1	Case Study 2	Case Study 3
Phase 1: Planning and Direction	Y/N	Y/N	Y/N
Phase 2: Collection	Y/N	Y/N	Y/N
Phase 3: Processing and Exploitation	Y/N	Y/N	Y/N
Phase 4: Analysis and Production	Y/N	Y/N	Y/N
Phase 5: Dissemination and Integration	Y/N	Y/N	Y/N
Phase 6: Evaluation and Feedback	Y/N	Y/N	Y/N

Source: Created by author using information from Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

After analyzing and reporting on each of the variables within each individual case study, the researcher conducts a cross case synthesis. The purpose of this cross-case synthesis is for two purposes: 1) identify which unintended bias indicators exist in which JIP phase, and 2) identify which participatory design best practice mitigation strategy should be employed against which unintended biases. To identify which unintended biases exist in which JIP phase, the researcher cross references Table 1 (presence of unintended bias indicators) and Table 3 (presence of JIP related tasks); Table 4 reflects the overall results. To identify which participatory design best practice mitigation strategy should be employed against which unintended biases, the researcher cross references Table 1 (presence of unintended bias indicators) and Table 2 (presence of participatory design strategies); Table 5 reflects the overall results. From Tables 4 and 5 the military intelligence community can see which unintended bias indicators may appear in which JIP phase as well as which participatory design best practice mitigation strategy to employ against which unintended bias.

Table 4. Template of Cross Case Synthesis of Variables from Tables 1 and 3

Bias Type	Phase 1: Planning and Direction	Phase 2: Collection	Phase 3: Processing and Exploitation	Phase 4: Analysis and Production	Phase 5: Dissemination and Integration	Phase 6: Evaluation and Feedback
Sampling	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N
Measurement	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N
Label	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N
Negative Set	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N
Framing Effect	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N
Sample Selection	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N
Confounding	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N
Design	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N
Sample Treatment	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N
Human Evaluation	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N
Test Dataset	Y/N	Y/N	Y/N	Y/N	Y/N	Y/N

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134; Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

Table 5. Template of Cross-Case Synthesis of Variables from Tables 1 and 2

Bias Type	Co-Design (CS#)	Collaboration (CS#)	Contribution (CS#)	Consultation (CS#)
Sampling	Y/N	Y/N	Y/N	Y/N
Measurement	Y/N	Y/N	Y/N	Y/N
Label	Y/N	Y/N	Y/N	Y/N
Negative Set	Y/N	Y/N	Y/N	Y/N
Framing Effect	Y/N	Y/N	Y/N	Y/N
Sample Selection	Y/N	Y/N	Y/N	Y/N
Confounding	Y/N	Y/N	Y/N	Y/N
Design	Y/N	Y/N	Y/N	Y/N
Sample Treatment	Y/N	Y/N	Y/N	Y/N
Human Evaluation	Y/N	Y/N	Y/N	Y/N
Test Dataset	Y/N	Y/N	Y/N	Y/N

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134; Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, London, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf.

Summary

This chapter discusses the qualitative, multiple case study methodology used for this study. As aforementioned, the various unintended bias indicators and participatory design best practices mitigation strategies may not be present, nor applicable, in every situation. This indicates that a qualitative, vice a quantitative, study is more appropriate as the study's recommendations and findings will likely be more nuanced in nature.[152] In addition, this study analyzes commercial industry participatory design best practices, for

[152] Yin, *Case Study Research and Applications,* 3-23.

mitigating unintended biases, to better inform the military intelligence community of how they can utilize them.

CHAPTER 4

ANALYSIS

Introduction

This chapter covers all the individual case study analysis as well as the cross-case synthesis. While the results of the individual case studies correlate to the secondary research questions, the results of the cross-case synthesis focus on answering the primary research question. This chapter covers the results for each individual case study analysis, the overall results of the study's variable analysis, and the results of the cross-case synthesis.

Case Study 1

Background

"412 Food Rescue" is a non-profit organization that works with "food retailers, volunteer drivers, and nonprofit organizations to connect surplus food with individuals and families who are experiencing food insecurity."[153] While founded in 2015, the organization realized it struggled with fairly distributing food to various nonprofit organizations in need. As such, it decided to create a technological system to assist with this process. In 2016, the organization worked with Carnegie Mellon University researchers using participatory design to co-create an algorithm for this system. Carnegie

[153] "What We Do," 412 Food Rescue, accessed April 11, 2022, https://412foodrescue.org/about-us/what-we-do/.

Mellon and Food Rescue 412 recorded part of this co-creation process in a joint study.[154] The study focused predominately on what the organization, donors, volunteers, and individual recipients considered "fair" to prevent unintended bias in "allocation" of food resources. [155] These efforts informed the creation of the 412 Food Rescue mobile application.[156] The original application is still active but it only serves a couple area codes in and around Pittsburg, Pennsylvania; the location of the organization's headquarters.[157] Eventually, this resulted in the subsequent creation of the "Food Rescue Hero" mobile application which has an expanded geographical range to correspond with 412 Food Rescue's operational broadening across the country.[158] Food Rescue Hero reportedly "employs machine-learning algorithms to efficiently match available food to a beneficiary organization's particular needs."[159] These applications serve the same

---

[154] Min Kyung Lee, Ji Tae Kim, and Leah Lizarondo, "A Human-Centered Approach to Algorithmic Services: Considerations for Fair and Motivating Smart Community Service Management that Allocates Donations to Non-Profit Organizations," (paper presented at the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, May 6-11, 2017), 3365-3376, https://dl.acm.org/doi/pdf/10.1145/3025453.3025884.

[155] Ibid., 3365.

[156] Google, "412 Food Rescue," Google Play, accessed April 11, 2022, https://play.google.com/store/apps/details?id=com.fouronetwo.foodrescue&referrer=utm_source%3Dinstagramadsummer18%26utm_medium%3DSocial; 412 Food Rescue, "412 Food Rescue," accessed April 11, 2022, https://412foodrescue.org/.

[157] Google, "412 Food Rescue"; "412 Food Rescue," 412 Food Rescue.

[158] Google, "Food Rescue Hero," Google Play, accessed April 11, 2022, https://play.google.com/store/apps/details?id=org.foodrescuehero.app; Kate Kelly, "How One App Saved Over 40 Million Pounds of Food from the Landfill," *Tech Soup* (blog), *Tech Soup*, March 15, 2021, https://blog.techsoup.org/posts/how-one-app-saved-over-40-million-pounds-of-food-from-the-landfill.

purpose in that they assist 412 Food Rescue with collecting information from donors and processing this information to identify local nonprofits in need. Once the application's algorithm matches the donor to a recipient, the algorithm disseminates a notification to local volunteers to complete the pickup and delivery. When the volunteer completes the delivery, the action is complete.[160]

Presence of Variables in Case Study 1

Table 6. Case Study 1 Variable 1 Results

Bias Group	Bias Type	Case Study 1:	Case Study 2:	Case Study 3:
Data Creation	Sampling	N	Y	Y
	Measurement	N	N	Y
	Label	N	Y	Y
	Negative Set	N	Y	Y
Problem Formulation	Framing Effect	Y	N	N
Data Analysis	Sample Selection	N	Y	Y
	Confounding	N	N	N
	Design	Y	N	N
Evaluation & Validation	Sample Treatment	N	N	N
	Human Evaluation	Y	N	N
	Validation and Test Dataset	N	Y	Y

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134.

[159] AI.Business, "Can Artificial Intelligence Reduce Food Waste?" AI, January 25, 2017, https://ai.business/2017/01/25/can-artificial-intelligence-reduce-food-waste/.

[160] "Food Rescue Hero," 412 Food Rescue, accessed April 11, 2022, https://412foodrescue.org/programs/foodrescuehero/.

This is a unique case study as it focuses on initial stages of creating an algorithm, presumably for an ML system; therefore, the analysis for variable 1 stems from what the designers identified as a potential issue or bias when building the system.[161] As reflected in Table 9, the study identifies the presence of three ML biases within the case study: 1) framing effect, 2) design, and 3) human evaluation. The study finds a presence of these three biases in the primary source material, notably the creator's joint study.[162]

An initial keyword search of the various bias sub-variables helps focus the analysis. The study identifies the presence of the framing effect bias as the ML system designers struggled with the complexity of selecting successful "allocation models."[163] The stakeholders from the joint study had differing views as to what was fair. Subsequently, this caused issues for the ML system designers to determine what constituted a fair allocation model. The three proposed models were based on "efficiency-, equality-, and equity-centric allocations."[164] While the designers did not specify the model they chose in their study's conclusion, all of their models had elements of fairness and unfairness according to the different stakeholder groups.[165]

---

[161] Lee, Kim, and Lizarondo, "A Human-Centered Approach to Algorithmic Services," 3365-3376.

[162] Ibid.

[163] Ibid., 3368.

[164] Ibid.

[165] Ibid., 3365-3376.

As for design bias, this bias includes the sub-types of ranking and presentation biases.[166] The study identifies the presence of ranking bias because stakeholders mentioned "concerns about ranking one organization over another…" when discussing advantages and disadvantages of the different allocation models.[167] Notably, one stakeholder mentioned the complexity in how to *"*measure the need;" however, upon review of the content the researcher interprets this as not an indication of measurement bias as the stakeholder mentioned the term "measure" in reference to bias in ranking organizations.[168]

Regardless of which model the designers chose to implement, the description of how the ML system operates also indicates there is some form of a ranking or presentation mechanism. One description specifies that "when there is a donation request, an algorithm will recommend recipient organizations and…the manager selects one or confirms…."[169] While the manager ultimately chooses the recipient organization, he or she chooses from the recommendations provided - ranked or presented - by the ML system.[170]

The last detected bias is human evaluation. As part of the joint study, the designers discuss the "need to understand stakeholders' unintentional biases that might

---

[166] Chander and Srinivasan, "Biases in AI Systems," 54.

[167] Lee, Kim, and Lizarondo, "A Human-Centered Approach to Algorithmic Services," 3370.

[168] Ibid.

[169] Ibid., 3368.

[170] Ibid.

make algorithmic systems function unfairly."[171] They referred to how volunteers may or may not sign up to assist if they feel the delivery location is in an unsafe area, leading to a bias related to the locations in which volunteers serve.[172] While other biases may be present in the ML system, these three biases were the only ones this study identifies in the case study based upon the collected material.

Table 7. Case Study 1 Variable 2 Results

Participatory Design Strategy	Case Study 1	Case Study 2	Case Study 3
Co-Creation	Y	N	N
Collaboration	Y	N	N
Contribution	N	Y	N
Consultation	N	N	Y

Source: Created by author using information from Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, London, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf.

[171] Ibid., 3373.

[172] Lee, Kim, and Lizarondo, "A Human-Centered Approach to Algorithmic Services," 3373.

The next variable for analysis is the presence of a participatory design best practice mitigation strategy. The 412 Food Rescue and Carnegie Mellon joint study specified that they wanted to use a "human-centered approach."[173] The study identifies the presence of two strategies, co-creation and collaboration, which are annotated in Table 10. As per the definition of co-creation, "external stakeholders help to initiate and oversee the project as well as collaborating throughout the AI pipeline."[174]

The study specifically recognizes this is a co-creation strategy as 412 Food Rescue "strategically decided to build an algorithmic system" which included an ML system.[175] Their organization worked directly with Carnegie Mellon researchers from the inception of the ML system. A technical team did not invite 412 Food Rescue to help improve a portion of an existing ML system; instead, 412 Food Rescue decided it wanted a technological platform to improve its operations and worked with external technical experts to design its ML system. As indicator of both co-creation and collaboration strategies, 412 Food Rescue and Carnegie Mellon researchers worked together in the design process of the ML system.[176] This demonstrated how technical and non-technical experts valued each other's feedback and incorporated it to build the ML system.[177] A similar study, which analyzed a follow-on joint study between Carnegie Mellon,

---

[173] Ibid., 3365.

[174] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 11.

[175] Lee, Kim, and Lizarondo, "A Human-Centered Approach to Algorithmic Services," 3367.

[176] Ibid., 3367-3368, 3372.

[177] Ibid., 3365-3376.

University of Texas at Austin, and 412 Food rescue, also corroborates that there is a presence of the collaboration participatory design mitigation strategy.[178]

Table 8. Case Study 1 Variable 3 Results

JIP Phases	Case Study 1	Case Study 2	Case Study 3
Phase 1: Planning and Direction	N	N	N
Phase 2: Collection	Y	Y	Y
Phase 3: Processing and Exploitation	Y	Y	Y
Phase 4: Analysis and Production	N	N	N
Phase 5: Dissemination and Integration	Y	N	N
Phase 6: Evaluation and Feedback	N	N	N

Source: Created by author using information from Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

The last variable focuses on the presence of JIP related tasks within the ML system. For this case study, the analyzed ML system is the Food Rescue Hero application.[179] (The original 412 Food Rescue application is presumably the same as the

[178] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 35.

[179] Apple, "Food Rescue Hero," App Store Preview, accessed April 11, 2022, https://apps.apple.com/us/app/food-rescue-hero/id1518660483.

newer application, except that it only operates in a specific area).[180] The study identifies JIP related tasks that fall within phases two, three, and five of the JIP, but does not identify the presence of JIP related tasks for phases one, four, and six. The researcher was unable to find any sources that indicated a presence of phase one JIP related tasks, planning and direction. The researcher's understanding is that the application is merely executing its tasks in accordance with the original plan and direction of its human designers. To access the mobile application, a user must sign up to be a volunteer by providing information in the application. This involves a phase two JIP related task of gathering information from the user.[181]

Next the study identifies that the application can complete phase three JIP related tasks of processing and exploiting as it is responsible for matching "food donations to the appropriate nonprofits" and providing these recommended matches to the Food Rescue manager for final approval.[182] Since the Food Rescue manager makes the decision of confirming or altering the match, the researcher interprets the ML system as not capable of executing phase four JIP related tasks of analysis and production as a human is ultimately making the analytical judgment.[183] As for phase five JIP related tasks, the application reportedly "coordinates a last-mile transportation network of volunteers" by

[180] Google, "412 Food Rescue."

[181] Apple, "Food Rescue Hero."

[182] 412 Food Rescue, "Food Rescue Hero"; Lee, Kim, and Lizarondo, "A Human-Centered Approach to Algorithmic Services," 3368.

[183] Lee, Kim, and Lizarondo, "A Human-Centered Approach to Algorithmic Services," 3373.

sending "out a call to volunteers" through the application.[184] While the ML system supposedly "tracks data and analytics," the researcher has not found any publicly available information as to how the application completes this task.[185] In addition, the researcher interprets this action as not an assessment of how the ML system is operating. Instead, it may merely be gathering and presenting information for a human to make the determination of the ML system's effectiveness. It is possible that the application can complete phase six JIP related tasks, evaluation and feedback, but additional information is needed to confirm.[186]

Case Study 2

Background

Despite obstacles, Google continues to pursue the development of its imagery ML systems. The Google photos application is one such ML system that allows users to store and find their photos with "Google's powerful image search abilities."[187] In 2015, a user identified an unintended bias in Google's photos application. Several news articles reported the incident which focused on the photos application misclassifying and

[184] "Food Rescue Hero," 412 Food Rescue; Lee, Kim, and Lizarondo, "A Human-Centered Approach to Algorithmic Services," 3368; Google, "412 Food Rescue"; Kelly, "How One App Saved Over 40 Million Pounds of Food from the Landfill."

[185] 412 Food Rescue, "Food Rescue Hero."

[186] 412 Food Rescue, "Food Rescue Hero."

[187] E. Justin Swanson, "Upload the Pictures, and Let Google Photos Do the Rest," *The New York Times*, June 3, 2015, https://www.nytimes.com/2015/06/04/technology/personaltech/upload-the-pictures-and-let-google-photos-do-the-rest.html.

mislabeling an individual as an animal.[188] Google quickly apologized and stated, "there is still clearly a lot of work to do with automatic image labeling, and we're looking at how we can prevent these types of mistakes from happening in the future."[189] Other users reported unintended bias in the photos application, such as "a picture of a friend's bloody elbow, injured while skateboarding, was labeled 'food.'"[190]

The following year, in 2016, Google implemented a participatory design best practice mitigation strategy of contribution by releasing its "Crowdsource" mobile application to the public.[191] This application allows public users to complete a variety of tasks, to include uploading photos and labeling the items in the photos.[192] With this Crowdsource application, Google built the Open Images Extended - Crowdsourced

---

[188] BBC, "Google Apologises for Photo App's Racist Blunder," *BBC News*, July 1, 2015, https://www.bbc.com/news/technology-33347866; Jessica Guynn, "Google Photos Labeled Black People 'Gorillas'," *USA Today*, July 1, 2015, https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/.

[189] Amanda Schupak, "Google Apologizes for mis-Tagging Photos of African Americans," *CBS News*, July 1, 2015, https://www.cbsnews.com/news/google-photos-labeled-pics-of-african-americans-as-gorillas/.

[190] Conor Dougherty, "Google Photos Mistakenly Labels Black People 'Gorillas'," *Business, Innovation, Technology, Society* (blog), *The New York Times*, July 1, 2015, https://bits.blogs.nytimes.com/2015/07/01/google-photos-mistakenly-labels-black-people-gorillas/.

[191] Google, "Crowdsource," Google Play, accessed April 11, 2022, https://play.google.com/store/apps/details?id=com.google.android.apps.village.boond.

[192] Google, "Crowdsource"; Supheakmungkol Sarin, Knot Pipatsrisawat, Khiem Pham, Anurag Batra, and Lu´ıs Valente, "Crowdsource by Google: A Platform for Collecting Inclusive and Representative Machine Learning Data," (paper presented in the Work in Progress and Demo Track at the seventh AAAI Conference on Human Computation and Crowdsourcing 2019, Skamania Lodge, WA, October 28-30, 2019), https://storage.googleapis.com/pub-tools-public-publication-data/pdf/55f12b7d77b32432b36970ac4e9ee57cb546ca7f.pdf.

dataset, which "aims to provide researchers and developers with more geo-diverse images for model training and testing."[193] This is an extension to the Open Images dataset that Google "professional annotators" maintain and make available to the public.[194]

Google has not explicitly stated that the creation of this crowdsource application and associated dataset is a mitigation response to the aforementioned incidents; however, Google recognizes that its imagery training dataset lacks "geographical and demographic diversity," especially from communities in "Asia, Africa, and Latin America."[195] According to its Google Data Card, Google created the Open Images Extended – Crowdsourced dataset to help train and test imagery ML systems to "identify objects or context of photos visually…" and "find objects, plants, animals, etc. through search in Photos or Image Search."[196] The study is not able to determine if the Google Data Card is directly referring specifically to Google's "Photos or Image Search."[197] Google has mentioned that the "data gathered through the app is used across nearly all of Google's AI products, from identifying images using Google Lens, to organizing pictures in

---

[193] Google, "Open Images Extended - Crowdsourced," Google Research, 2018, accessed April 11, 2022, https://research.google/tools/datasets/open-images-extended-crowdsourced/; Kyle Wiggers, "Google's Inclusive Images Competition Spurs Development of Less Biased Image Classification AI," Venture Beat, December 2, 2018, https://venturebeat.com/2018/12/02/googles-inclusive-images-competition-spurs-development-of-less-biased-image-classification-ai/.

[194] Google, "Overview of Open Images V6," Open Images Dataset V6, accessed April 11, 2022, https://storage.googleapis.com/openimages/web/factsfigures.html.

[195] Google, "Open Images Extended - Crowdsourced," Google Research.

[196] Google, "Open Images Extended - Crowdsourced," Google LLC, accessed April 11, 2022, https://research.google/static/documents/datasets/open-images-extended-crowdsourced.pdf.

[197] Ibid.

Photos…."[198] It also posted in its blog that with Crowdsource application users can "help train Google's ML models and AI…."[199] This indicates that Google uses the crowdsourced data to train its ML systems, but the study is unable to verify at this time if Google is using the Open Images Extended – Crowdsourced dataset in its entirety for testing and training of its non-open proprietary ML systems.

Presence of Variables in Case Study 2

Table 9. Case Study 2 Variable 1 Results

Bias Group	Bias Type	Case Study 1:	Case Study 2:	Case Study 3:
Data Creation	Sampling	N	**Y**	Y
	Measurement	N	**N**	Y
	Label	N	**Y**	Y
	Negative Set	N	**Y**	Y
Problem Formulation	Framing Effect	Y	**N**	N
Data Analysis	Sample Selection	N	**Y**	Y
	Confounding	N	**N**	N
	Design	Y	**N**	N
Evaluation & Validation	Sample Treatment	N	**N**	N
	Human Evaluation	Y	**N**	N
	Validation and Test Dataset	N	**Y**	Y

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134.

[198] Anurag Batra, "How to Help Make AI Systems More International and Inclusive with Google's Crowdsource App," Accelerate with Google, accessed April 11, 2022, https://accelerate.withgoogle.com/stories/how-to-help-make-AI-systems-more-international-and-inclusive-with-googles-crowdsource-app.

[199] Team Crowdsource, "Welcome to the Crowdsource by Google Blog!" *Crowdsource by Google* (blog), *Google*, May 2020, https://crowdsource.google.com/about/blog/welcome-to-the-crowdsource-blog/.

Unlike the first case study, this case study involves the detection of unintended bias within existing ML systems, specifically Google's imagery ML systems. Of note, this includes the Google Photos application. As reflected in Table 12 the study identifies the presence of five ML biases within the case study: 1) sampling, 2) label, 3) negative set, 4) sample selection, and 6) validation and test dataset. The study detects sampling bias from Google's own admission that its training dataset is insufficiently diverse. As most of the images in Google's training dataset are reportedly samples from the "United States and Western Europe" this indicates that the training data is "skewed" towards certain cultures.[200]

The next two unintended biases, label and negative set, are also related to the training dataset for Google's imagery ML systems. Google's Photos application made errors identifying objects in images, and subsequently, mislabeling them. This indicates that the ML system also had difficulties in ruling out what the images were not, which is negative set bias. In these same instances, the ML demonstrated a label bias when it mislabeled objects in an image.[201] Sampling selection relates to the first three unintended biases. When Google announced it had insufficient diverse images for its training

[200] Tom Simonite, "Google Turns to Users to Improve its AI Chops Outside the US," *Wired*, April 5, 2018, https://www.wired.com/story/google-turns-to-users-to-improve-its-ai-chops-outside-the-us/.

[201] BBC, "Google Apologises for Photo App's Racist Blunder"; Guynn, "Google Photos Labeled Black People 'Gorillas',"; Dougherty, "Google Photos Mistakenly Labels Black People 'Gorillas',"; Simonite, "Google Turns to Users to Improve its AI Chops Outside the US."

datasets, it indicated a sampling selection bias.[202] The last bias that the study detects is the test dataset bias. With the presence of the first three biases, this indicates that the validation and test dataset for the ML system suffers from "biases associated with dataset-creation stage."[203] The study could not confirm the presence of the other unintended biases due to lack of sufficient detail in the collected, publicly available information.

For this case study, the researcher detects the presence of the contribution participatory design best practice mitigation strategy. Contribution refers to "participation that is usually time-limited to one stage of the AI development pipeline and involves external stakeholders completing one of the tasks that is necessary to AI development…"[204] In addition, "common methods include both targeted and open crowdsourcing."[205] As such, the researcher identifies its presence as Google openly stated it was using crowdsourcing to diversify its training dataset.[206] It even named its

---

[202] Anurag Batra and Parker Barnes, "Adding Diversity to Images with Open Images Extended," *Google AI Blog* (blog), *Google Research*, December 7, 2018, https://ai.googleblog.com/2018/12/adding-diversity-to-images-with-open.html; Tulsee Doshi, "Introducing the Inclusive Images Competition," *Google AI Blog* (blog), *Google Research*, September 6, 2018, https://ai.googleblog.com/2018/09/introducing-inclusive-images-competition.html.

[203] Chander and Srinivasan, "Biases in AI Systems," 55.

[204] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 12.

[205] Ibid.

[206] Anurag Batra, "How to Help Make AI Systems More International and Inclusive with Google's Crowdsource App"; Simonite, "Google Turns to Users to Improve its AI Chops Outside the US"; Sarin et al., "Crowdsource by Google," 1-3.

application the "Crowdsource" app.[207] While "time-limited" seems to imply a shorter time duration, Google has been using its Crowdsource app since 2016 to gather imagery related data; this suggests that Google still finds a need to continue the crowdsourcing to improve its test dataset, and subsequently, its ML systems.[208]

Table 10. Case Study 2 Variable 2 Results

Participatory Design Strategy	Case Study 1	Case Study 2	Case Study 3
Co-Creation	Y	N	N
Collaboration	Y	N	N
Contribution	N	Y	N
Consultation	N	N	Y

Source: Created by author using information from Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, London, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf.

[207] Google, "Crowdsource"; Google, "How Crowdsource Works," Crowdsource by Google, accessed April 13, 2022, https://crowdsource.google.com/about/how-it-works/.

[208] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 12.

Table 11. Case Study 2 Variable 3 Results

JIP Phases	Case Study 1	Case Study 2	Case Study 3
Phase 1: Planning and Direction	N	N	N
Phase 2: Collection	Y	Y	Y
Phase 3: Processing and Exploitation	Y	Y	Y
Phase 4: Analysis and Production	N	N	N
Phase 5: Dissemination and Integration	Y	N	N
Phase 6: Evaluation and Feedback	N	N	N

Source: Created by author using information from Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

Within the Google imagery ML systems, the study identifies JIP related tasks from two phases, collection as well as processing and exploitation. For specificity, the study refers to the Google Photos application and the tasks it completes. The Photos application completes collection related tasks regarding the "acquisition of data" as it gathers and stores information from users when they upload their images.[209] In addition, Google advertises that uploaded "photos are automatically organized and searchable" which indicates that "raw collected data is converted into forms that can be readily

[209] CJCS, JP 2-0, I-15; Google, "The Home for your Memories," Google Photos, accessed April 13, 2022, https://www.google.com/photos/about/; Anil Sabharwal, "Picture This: A Fresh Approach to Photos," *The Keyword* (blog), *Google*, May 28, 2015, https://blog.google/products/photos/picture-this-fresh-approach-to-photos/.

used…" by the consumer.[210] This capability indicates that the ML system is capable of completing processing and exploitation related tasks. Besides these tasks, the study does not identify the presence of any other JIP related tasks.

Case Study 3

Background

In 2017, a Massachusetts Institute of Technology (MIT) researcher and a Microsoft researcher conducted a study which involved a user audit of IBM, Microsoft, and Face++ facial recognition systems.[211] The MIT researcher had detected a bias in some facial recognition systems, regarding skin type and gender, and wanted to test how accurate the different commercial companies' ML systems were.[212] For the study, they built the "Pilot Parliaments Benchmark" (PPB) which is a "new dataset with more balanced skin type and gender representations."[213] They used the PPB to test the accuracy

---

[210] Google, "The Home for your Memories"; Shimrit Ben-Yair, "Your Photos, Your Memories, Your Way," *The Keyword* (blog), *Google*, May 18, 2021, https://blog.google/products/photos/new-memories-features-look-back/; CJCS, JP 2-0, I-15.

[211] Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," Proceedings of *Machine Learning Research* 81 (2018): 1-15 (Conference on Fairness, Accountability, and Transparency, New York University, New York City, February 23-24, 2018), http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

[212] Steve Lohr, "Facial Recognition is Accurate, if You're a White Guy*," The New York Times*, February 9, 2018, https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html?action=click&module=RelatedLinks&pgtype=Article.

[213] Buolamwini and Gebru, "Gender Shades," 4; "Gender Shades," MIT Media Lab, 2018, accessed April 15, 2022, http://gendershades.org/overview.html.

of the three facial recognition systems and found inaccuracies in all three systems.[214] In the study's conclusion it stated, "We found that all classifiers performed best for lighter individuals and males overall. The classifiers performed worst for darker females."[215]

Upon completion of the study, they shared the results with all three companies.[216] IBM and Microsoft responded to the researchers and provided respective updates to their systems.[217] Notably, IBM had already been working on improvements in its Watson Visual Recognition system, but also used input from the study to further test the system.[218] Face++ also published an updated version of its facial recognition system.[219] In addition to notifying the respective companies, the authors published as well as engaged with the media and the public to disseminate the study's findings.[220] This led to

---

[214] Buolamwini and Gebru, "Gender Shades," 12.

[215] Ibid.

[216] "Gender Shades," MIT Media Lab.

[217] "Gender Shades," MIT Media Lab; James Vincent, "IBM Hopes to Fight Bias in Facial Recognition with New Diverse Dataset," *The Verge*, June 27, 2018, https://www.theverge.com/2018/6/27/17509400/facial-recognition-bias-ibm-data-training; IBM Research Editorial Staff, "IBM to Release World's Largest Annotation Dataset for Studying Bias in Facial Analysis," *IBM Research Blog* (blog), *IBM*, June 27, 2018, https://www.ibm.com/blogs/research/2018/06/ai-facial-analytics/; John Roach, "Microsoft Improves Facial Recognition Technology to Perform Well across All Skin Tones, Genders," *The AI Blog* (blog), *Microsoft*, June 26, 2018, https://blogs.microsoft.com/ai/gender-skin-tone-facial-recognition-improvement/.

[218] Gender Shades, "IBM Response to 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," MIT Media Lab, January 2018, 1-4, http://gendershades.org/docs/ibm.pdf.

[219] MEGVII, "Notice: Newer Version of Face Detect API," *Face++* (blog), June 27, 2018, https://www.faceplusplus.com/blog/article/newer-version-face-detect-api/?cate=product_news.

[220] "Gender Shades." MIT Media Lab.

the second study which involved a re-audit on IBM, Microsoft, and Face++ facial recognition systems as well as a new audit on Amazon and Kairos facial recognition systems.[221] In the second study, the authors discovered there were less inaccuracies in IBM, Microsoft, and Face++ facial recognition systems, in comparison to the first study's findings as well as in contrast with the higher inaccuracies it found in Amazon and Kairos.[222] Of the audited companies, Amazon expressed concern and described the study's findings as "misleading;" however, Amazon announced it was interested in further discussions regarding the findings.[223] With the publication of these studies, the media have reported public and government concern regarding the use of facial

[221] Inioluwa Deborah Raji and Joy Buolamwini, "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products," *Proceedings of the 2018 Association for the Advancement of Artificial Intelligence (AAAI/ACM)* (2019): 1-7 (Conference on AI, Ethics, and Society (AIES '19), Honolulu, Hawaii, January 27-28, 2019), https://www.thetalkingmachines.com/sites/default/files/2019-02/aies-19_paper_223.pdf; Natasha Singer, "Amazon is Pushing Facial Technology that a Study Says Could be Biased," *The New York Times*, January 24, 2019, https://www.nytimes.com/2019/01/24/technology/amazon-facial-technology-study.html.

[222] Raji and Buolamwini, "Actionable Auditing," 4.

[223] Matt Wood, "Thoughts on Recent Research Paper and Associated Article on Amazon Rekognition," *AWS Machine Learning Blog* (blog), *AWS*, January 26, 2019, https://aws.amazon.com/blogs/machine-learning/thoughts-on-recent-research-paper-and-associated-article-on-amazon-rekognition/; Raji and Buolamwini, "Actionable Auditing," 6; Joy Buolamwini, "Re: Audit of Amazon Rekognition Uncovers Gender and Skin-Type Disparities," (e-mail to Founder and Chief Executive Officer of Amazon, Inc., June 25, 2018), https://uploads.strikinglycdn.com/files/e286dfe0-763b-4433-9a4b-7ae610e2dba1/RekognitionGenderandSkinTypeDisparities-June25-Mr.%20Bezos.pdf.

recognition in certain capacities, specifically law enforcement.[224] In addition, IBM publicly voiced concerns about the use of facial recognition software in law enforcement and "has sunset its general purpose facial recognition and analysis software products."[225] Additionally, "Microsoft has announced that it will not sell facial recognition technology to police departments…."[226] Altogether the authors have implemented a consultation participatory design best practice mitigation strategy with their combination of the user audits and public engagement regarding facial recognition usage. This consultation not only motivated some of the audited companies to improve their ML systems but has also helped inspire Congress to introduce legislation on the subject.[227]

---

[224] Drew Harwell, "Federal Study Confirms Racial Bias of Many Facial-Recognition Systems, Casts Doubt on Their Expanding Use," *The Washington Post*, December 19, 2019, https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/; *Facial Recognition Technology (Part 1): Its Impact on Our Civil Rights and Liberties, Written Testimony of Joy Buolamwini*, U.S. House Committee on Oversight and Government Reform, 116th Cong., May 22, 2019, 1-20, https://docs.house.gov/meetings/GO/GO00/20190522/109521/HHRG-116-GO00-Wstate-BuolamwiniJ-20190522.pdf; Shira Ovide, "A Case for Banning Facial Recognition," *The New York Times*, updated August 1, 2021, https://www.nytimes.com/2020/06/09/technology/facial-recognition-software.html; Buolamwini, "When the Robot Doesn't See Dark Skin."

[225] IBM, "IBM CEO's Letter to Congress on Racial Justice Reform," *THINKPolicy Blog* (blog), June 8, 2020, https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/?mhsrc=ibmsearch_a&mhq=congress%20facial%20recognition; Khari Johnson, "IBM Walked Away from Facial Recognition. What about Amazon and Microsoft?" Venture Beat, June 10, 2020, https://venturebeat.com/2020/06/10/ibm-walked-away-from-facial-recognition-what-about-amazon-and-microsoft/.

[226] "What is Azure Face Service?" Microsoft, March 2, 2022, https://docs.microsoft.com/en-us/azure/cognitive-services/face/overview.

[227] *Facial Recognition Technology (Part 1)*; Facial Recognition and Biometric Technology Moratorium Act of 2021, S. Res. 117th Cong., 1st sess. (June 15, 2021): S.2052, https://www.congress.gov/bill/117th-congress/senate-bill/2052/all-info?r=1&s=1.

This study relies upon English speaking sources predominately originating from the two MIT academic studies, announcements from Microsoft, IBM, Amazon, and Kairos as well as the U.S. news. The study focuses on Microsoft, IBM, Amazon, and Kairos because primary information is difficult to find regarding Face++. This latter company is Chinese based, and has an English-speaking website; however, Face++ does not appear in other English-speaking sources, such as the media, as prevalently as the U.S. companies.[228] There might be more information available in Chinese sources.

Presence of Variables in Case Study 3

Table 12. Case Study 3 Variable 1 Results

Bias Group	Bias Type	Case Study 1:	Case Study 2:	Case Study 3:
Data Creation	Sampling	N	Y	**Y**
	Measurement	N	N	**Y**
	Label	N	Y	**Y**
	Negative Set	N	Y	**Y**
Problem Formulation	Framing Effect	Y	N	**N**
Data Analysis	Sample Selection	N	Y	**Y**
	Confounding	N	N	**N**
	Design	Y	N	**N**
Evaluation & Validation	Sample Treatment	N	N	**N**
	Human Evaluation	Y	N	**N**
	Validation and Test Dataset	N	Y	**Y**

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134.

[228] MEGVII, "Face++," accessed April 15, 2022, https://www.faceplusplus.com/.

In the third case study, the study identifies six overall unintended biases in the facial recognition ML systems. The unintended biases are as follows: 1) sampling, 2) measurement, 3) label, 4) negative set, 5) sample selection, and 6) validation and test dataset. As highlighted in the two academic studies, all the tested ML systems had varying levels of inaccuracies when tested against the PPB test dataset.[229] This served as an indication to the researcher that there is a presence of sampling bias in the various ML systems' original datasets as some of the systems had less inaccuracies with some skin tones and genders over others.[230] In the first audit, Microsoft, IBM, and Face++ "performed best for lighter individuals and males overall" with less inaccuracies in the second audit after they adjusted their respective systems, predominately the training datasets.[231] In the second audit, the authors included Amazon and Kairos which also tested "better on male faces than female faces…" and "on lighter faces than darker faces…."[232] Based upon these results, the ML systems' original training or test datasets were insufficiently diverse so some "instances" were more present than others.[233] While Microsoft did not declare it had a sampling bias in its training or test datasets when it published its updated ML system, it stated, "improving the performance of the gender

[229] Buolamwini and Gebru, "Gender Shades," 1-15; Raji and Buolamwini, "Actionable Auditing," 1-7.

[230] Buolamwini and Gebru, "Gender Shades," 1.

[231] Buolamwini and Gebru, "Gender Shades," 12; Raji and Buolamwini, "Actionable Auditing," 4.

[232] Raji and Buolamwini, "Actionable Auditing," 4.

[233] Chander and Srinivasan, "Biases in AI Systems," 48.

classifier in the Face API was mainly a technical challenge."[234] Microsoft also stressed the importance of "collecting more data that captures the diversity of our world…."[235]

Next the study detects a presence of measurement bias as annotated in the two MIT studies where the authors state the "potential physical limitations of the image quality and illumination of darker skinned subjects may be contributing to the higher error rate for that group…."[236] The authors took steps to ensure their PPB test dataset had "high image resolution, and the consistency of illumination and pose…," and the audited ML systems still had errors.[237] In IBM's response to the initial audit, it specified that it had "several ongoing projects to address dataset bias in facial analysis – including…. actors such as pose, illumination, resolution…."[238] As for the third bias, in the first audit the authors extensively discussed how they decided upon labels in which they settled on using darker versus lighter "skin type labels" and "the binary sex labels of female and male" for the "gender labels."[239] In both audits, the ML systems sometimes errored in appropriately identifying gender and skin types which indicates there was a labeling bias

---

[234] Roach, "Microsoft Improves Facial Recognition Technology to Perform Well across All Skin Tones, Genders."

[235] Ibid.

[236] Raji and Buolamwini, "Actionable Auditing," 5; Buolamwini and Gebru, "Gender Shades," 11-12.

[237] Buolamwini and Gebru, "Gender Shades," 12.

[238] Gender Shades, "IBM Response to 'Gender Shades'," 2.

[239] Buolamwini and Gebru, "Gender Shades," 6; Larry Hardesty, "Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems," *MIT News*, February 11, 2018, https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212.

in the ML systems.[240] In fact, IBM specifically discussed testing their updated ML system with new labels, similar to those in the first audit.[241] Due to the presence of the first three biases, the study identifies a negative set bias as the ML systems erroneously misidentified skin and gender types. This implies that the systems could not identify that an individual was of the opposite gender or had a different skin type in some images.[242]

The last two detected biases are closely related to the presence of the initial four biases. Due to the lack of diverse samples in the datasets, "the selection of individuals, groups, or data for analysis…are not representative of the population intended to be analyzed…"[243] As there appeared to be prevalence of "lighter-skinned…" samples in the datasets, the sample selection would also cause the ML systems to make inaccurate "correlations."[244] In IBM's response to the MIT study, it announced and published a "million-scale dataset of face images…to reduce sample selection bias."[245]

The last unintended bias is the validation and test dataset. Due to the errors detected from the MIT studies when the authors used a different validation and test

---

[240] Buolamwini and Gebru, "Gender Shades," 10; Raji and Buolamwini, "Actionable Auditing," 4.

[241] Ruchir Puri, "Mitigating Bias in AI Models," *IBM Research Blog* (blog), *IBM*, February 6, 2018, https://www.ibm.com/blogs/research/2018/02/mitigating-bias-ai-models/.

[242] Raji and Buolamwini, "Actionable Auditing," 4; Buolamwini and Gebru, "Gender Shades," 10.

[243] Chander and Srinivasan, "Biases in AI Systems," 52.

[244] Buolamwini and Gebru, "Gender Shades," 1; Chander and Srinivasan, "Biases in AI Systems," 52.

[245] Gender Shades, "IBM Response to 'Gender Shades'," 2.

dataset to audit the ML systems, this study detects this bias's presence in the various ML systems' original test and validation datasets. In response to the original audit, IBM stated that it had "been working towards substantially increasing the accuracy of its new Watson Visual recognition for facial analysis, which now uses different training data…."[246] IBM further decided to "evaluate IBM's new service in a manner consistent with their [Gender Shade's] study" with a dataset "very similar to the Pilot Parliaments Benchmark."[247]

Microsoft also announced that in their updated facial recognition system their technical team had "expanded and revised training and benchmark datasets, launched new data collection efforts to further improve the training data by focusing specifically on skin tone, gender and age, and improved the classifier to produce higher precision results" as well as "talked about different strategies to internally test our systems…."[248] Both companies specifically identified issues with their training datasets as well as expressed concerns about, or indicated additional measures regarding, testing which suggests the presence of bias within validation and test datasets.[249] As for confounding bias, the study was unable to confirm its presence in the case study. The second MIT study discussed that, "If a prediction is not produced (i.e. face not detected), we omit the

[246] Gender Shades, "IBM Response to 'Gender Shades'," 1.

[247] Puri, "Mitigating Bias in AI Models."

[248] Roach, "Microsoft Improves Facial Recognition Technology to Perform Well across All Skin Tones, Genders."

[249] Chander and Srinivasan, "Biases in AI Systems," 55.

result from our calculations."[250] This suggests some of the ML systems were unable to detect a face, but without proprietary information, the researcher is unable to determine if this is because the ML system is "not taking into account all the information in the data or if it misses the relevant relations between features and target outputs…."[251]

Table 13. Case Study 3 Variable 2 Results

Participatory Design Strategy	Case Study 1	Case Study 2	Case Study 3
Co-Creation	Y	N	**N**
Collaboration	Y	N	**N**
Contribution	N	Y	**N**
Consultation	N	N	**Y**

Source: Created by author using information from Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, London, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf.

For the third case study, the study identifies the presence of the consultation participatory design best practice mitigation strategy. The MIT studies were "outside of

[250] Raji and Buolamwini, "Actionable Auditing," 3.

[251] Chander and Srinivasan, "Biases in AI Systems," 52.

the core AI development process" of the respective companies.[252] In addition, "the goals of consultations are typically twofold; some initiatives seek to understand public attitudes for different applications of AI or regulatory policy, while others aim to increase data literacy and raise awareness of the impacts of AI systems."[253] The MIT studies helped inspire change in some of the audited companies and their respective ML systems as well as educate and engage with the public to motivate regulatory action.[254] The initial audit study was "accompanied by a video, summary visualizations and a website [which] further prompts public, academic and corporate audiences - technical and non-technical alike - to be exposed to the issue and respond."[255] The public, academia, media, and the government have referenced the study to discuss and conduct further research regarding the usage of facial recognition software, especially in law enforcement.[256] Even the audited companies, as well as others, have responded by either updating their ML systems, calling for regulatory action to restrict facial recognition in law enforcement, or

[252] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 11.

[253] Ibid., 11.

[254] Ibid., 11-12; Raji and Buolamwini, "Actionable Auditing," 6.

[255] Raji and Buolamwini, "Actionable Auditing," 2.

[256] Erik Learned-Miller, Vicente Ordonez, Jamie Morgenstern, and Joy Buolamwini, *Facial Recognition Technologies in the Wild: A Call For a Federal Office* (Cambridge, MA: Algorithmic Justice League, May 29, 2020), https://assets.website-files.com/5e027ca188c99e3515b404b7/5ed1145952bc185203f3d009_FRTsFederalOfficeMay2020.pdf; Harwell, "Federal Study Confirms Racial Bias of Many Facial-Recognition Systems, Casts Doubt on Their Expanding Use"; *Facial Recognition Technology (Part 1)*; Buolamwini, "When the Robot Doesn't See Dark Skin"; Khari Johnson, "Congress Moves toward Facial Recognition Regulation," Venture Beat, January 15, 2020, https://venturebeat.com/2020/01/15/congress-moves-toward-facial-recognition-regulation/.

no longer offering some of their facial recognition services.[257] While audits generally align with the co-creation category, this study concludes the case study actually falls within the consultation category due to the evidence previously stated.[258]

Table 14. Case Study 3 Variable 3 Results

JIP Phases	Case Study 1	Case Study 2	Case Study 3
Phase 1: Planning and Direction	N	N	N
Phase 2: Collection	Y	Y	Y
Phase 3: Processing and Exploitation	Y	Y	Y
Phase 4: Analysis and Production	N	N	N
Phase 5: Dissemination and Integration	Y	N	N
Phase 6: Evaluation and Feedback	N	N	N

Source: Created by author using information from Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

[257] IBM, "IBM CEO's Letter to Congress on Racial Justice Reform"; Jay Greene, "Microsoft Won't Sell Police Its Facial-Recognition Technology, Following Similar Moves by Amazon and IBM," *The Washington Post*, June 11, 2020, https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/; Jeffrey Dastin and Munsif Vengattil, "Microsoft Bans Face-Recognition Sales to Police as Big Tech Reacts to Protests," *Reuters*, June 11, 2020, https://www.reuters.com/article/us-microsoft-facial-recognition/microsoft-bans-face-recognition-sales-to-police-as-big-tech-reacts-to-protests-idUSKBN23I2T6; Brad Smith, "Finally, Progress on Regulating Facial Recognition," *Microsoft on the Issues* (blog), *Microsoft*, March 31, 2020, https://blogs.microsoft.com/on-the-issues/2020/03/31/washington-facial-recognition-legislation/.

[258] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 13-14.

Like the second case study, the study only finds the presence of JIP related tasks from two phases: collection and processing and exploitation. The ML systems could gather information as users take photos with or upload photos into them.[259] This indicates the ML systems complete phase two, collection, related tasks. As for processing and exploitation, all the systems are also designed to at least detect faces in images which indicates that "raw collected data is converted into forms that can be readily used…" by consumers.[260]

Case Study Analysis Results

The overall results of the individual case study analysis show a presence of all three variables in each case study. Regarding the first variable, unintended biases, the study identifies substantially different unintended biases in the first versus second and third case studies. The second and third case studies only have one difference between them which is the presence of measurement bias in the third case study. Table 15 displays the presence of unintended bias results for all three case studies. The second variable, presence of participatory design best practice mitigation strategies, varies across the case studies. The study annotates that the first case study has a presence of two participatory design mitigation strategies while the second and third case studies each have the presence of only one strategy. Table 16 displays the presence of participatory design best

[259] Microsoft, "Face API," Microsoft Azure, accessed April 15, 2022, https://azure.microsoft.com/en-us/services/cognitive-services/face/; Face++, "Detect API," Doc Center, accessed April 15, 2022, https://console.faceplusplus.com/documents/5679127; Amazon, "Detecting and Analyzing Faces," AWS, accessed April 15, 2022, https://docs.aws.amazon.com/rekognition/latest/dg/faces.html; Kairos, "API Reference," Developing with Kairos, accessed April 15, 2022, https://www.kairos.com/docs/api/.

[260] CJCS, JP 2-0, I-15.

practice mitigation strategies for all three case studies. The presence of the last variable, JIP related tasks, is fairly consistent across all three case studies, except that the first case study has tasks related to one additional phase. As the study does not include proprietary information, the study considers the entire application or system in each case study as the ML system. Table 17 displays the presence of JIP related tasks for all three case studies.

Table 15. Case Study Variable 1 Results

Bias Group	**Bias Type**	**Case Study 1:**	**Case Study 2:**	**Case Study 3:**
Data Creation	Sampling	N	Y	Y
	Measurement	N	N	Y
	Label	N	Y	Y
	Negative Set	N	Y	Y
Problem Formulation	Framing Effect	Y	N	N
Data Analysis	Sample Selection	N	Y	Y
	Confounding	N	N	N
	Design	Y	N	N
Evaluation & Validation	Sample Treatment	N	N	N
	Human Evaluation	Y	N	N
	Validation and Test Dataset	N	Y	Y

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134.

Table 16. Case Study Variable 2 Results

Participatory Design Strategy	Case Study 1	Case Study 2	Case Study 3
Co-Creation	Y	N	N
Collaboration	Y	N	N
Contribution	N	Y	N
Consultation	N	N	Y

Source: Created by author using information from Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, London, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf.

Table 17. Case Study Variable 3 Results

JIP Phases	Case Study 1	Case Study 2	Case Study 3
Phase 1: Planning and Direction	N	N	N
Phase 2: Collection	Y	Y	Y
Phase 3: Processing and Exploitation	Y	Y	Y
Phase 4: Analysis and Production	N	N	N
Phase 5: Dissemination and Integration	Y	N	N
Phase 6: Evaluation and Feedback	N	N	N

Source: Created by author using information from Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

Cross-Case Synthesis Results

Upon completion of the individual case study analyses, the researcher completes the cross-case synthesis. The study includes two approaches for the cross-case synthesis. The first approach involves cross-referencing the presence of unintended bias and JIP related task results within each individual case study. The researcher creates a chart for each case study that shows the results of the cross-referencing and overlays them to create an overall cross-reference table for the cross-case synthesis. The second approach focuses on cross-referencing the presence of unintended biases with the implemented participatory design best practice mitigation strategy within each individual case study. Like the first approach, the researcher creates a cross-reference chart of these variables for each individual case study and overlays them to create a second overall cross-reference table for the cross-case synthesis.

Cross Referenced Results Between Unintended Biases and JIP Phases

Table 18. Overall Cross Case Synthesis Results of Variables 1 and 3

Bias Type	Phase 1: Planning and Direction	Phase 2: Collection	Phase 3: Processing and Exploitation	Phase 4: Analysis and Production	Phase 5: Dissemination and Integration	Phase 6: Evaluation and Feedback
Sampling	N	Y	Y	N	N	N
Measurement	N	Y	Y	N	N	N
Label	N	Y	Y	N	N	N
Negative Set	N	Y	Y	N	N	N
Framing Effect	N	Y	Y	N	Y	N
Sample Selection	N	Y	Y	N	N	N
Confounding	N	N	N	N	N	N
Design	N	Y	Y	N	Y	N
Sample Treatment	N	N	N	N	N	N
Human Evaluation	N	Y	Y	N	Y	N
Test Dataset	N	Y	Y	N	N	N

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134; Chairman of the Joint Chiefs of Staff, Joint Publication 2-0, *Joint Intelligence* (Washington, DC: Joint Chiefs of Staff, October 2013), I-6, figure I-3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

After compiling all the cross-referencing tables of the first and third variables, the study identifies generally which unintended biases may appear in which JIP phases. The study displays these results in Table 18. As seen in the table, approximately nine unintended biases generally appear in the second and third JIP phases. The study only identifies three unintended biases in the fifth phase from the cross-case synthesis.

Cross Referenced Results Between Unintended Biases
and Participatory Design Strategies

Table 19. Overall Cross-Case Synthesis Results of Variables 1 and 2

Bias Type	Co-Design (CS1)	Collaboration (CS1)	Contribution (CS2)	Consultation (CS3)
Sampling	N	N	Y	Y
Measurement	N	N	N	Y
Label	N	N	Y	Y
Negative Set	N	N	Y	Y
Framing Effect	Y	Y	N	N
Sample Selection	N	N	Y	Y
Confounding	N	N	N	N
Design	Y	Y	N	N
Sample Treatment	N	N	N	N
Human Evaluation	Y	Y	N	N
Test Dataset	N	N	Y	Y

Source: Created by author using information from Ajay Chander and Ramya Srinivasan, "Biases in AI Systems," *ACM Queue* 19, no. 2 (2021): 56, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134; Aleks Berditchevskaia, Eirini Malliaraki, and Kathy Peach, "Participatory AI for Humanitarian Innovation: A Briefing Paper," (Nesta, London, UK, 2021), 11, https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf.

Like the first cross-case synthesis approach, the study compiles the three cross-referencing charts from each case study and overlays them to produce the overall results of which participatory design best practices generally mitigate which unintended biases. Table 19 above shows the results of the cross-case synthesis. Co-Design and Collaboration generally mitigate three biases, which are framing effect, design, and human evaluation. Contribution generally mitigates five biases which include sampling,

label, negative set, sample selection, and test dataset biases. The last participatory design mitigation strategy, consultation, generally mitigates six unintended biases, which are sampling, measurement, label, negative set, sample selection, and test dataset biases.

Summary

This chapter covered the data analysis results from the study's individual case studies as well as the cross-case synthesis. The study identified the presence of all three specified variables in all three case studies as well as identifies the presence of the variables' sub-variables. In addition, the study found that the various cross-referenced variables can give a generalized understanding of which unintended biases appear in which JIP phases as well as which best practice can mitigate those unintended biases. The next chapter covers the conclusions from this analysis as well as recommendations for additional research and implementation.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Introduction

This chapter covers the conclusions and recommendations of the study. It draws upon the analysis specified in the previous chapter as support. Due to the nature of this qualitative multiple case study and cross case synthesis, the final conclusions are generalized in nature. In addition, the study outlines current recommendations for improving the study itself and conducting a follow-on or repeat study. There are also recommendations for areas of future exploration and strategies for implementation within the military. This chapter first discusses the study's conclusions, and then the recommendations.

Conclusions

This study has several conclusions. This first section discusses these conclusions from the case study variable analysis, the cross-case synthesis, and the overall study. The study will cover these conclusions in the following order: 1) case study variable conclusions, 2) cross-case synthesis conclusions, and 3) research question conclusions.

Case Study Variable Conclusions

There is a presence of all three variables within each case study, despite their differences. The first case study involves using co-creation to mitigate unintended bias while designing an algorithm that executes tasks related to the first JIP phase (collection), second JIP phase (processing and exploitation), and fifth JIP phase (integration and dissemination). The second case study centers on using contribution, specifically

crowdsourcing, to mitigate unintended bias within datasets for ML systems that execute tasks related to the second JIP phase (collection) and third JIP phase (processing and exploitation). The third case study involves a consultation to mitigate unintended bias within datasets for ML systems that execute tasks related to the second JIP phase (collection) and third JIP phase (processing and exploitation).

In addition to variable presence within each case study, the study also identifies several overall trends from the case studies. The first case study generally focuses on mitigating unintended bias within an algorithm while the second and third case studies generally focus on mitigating unintended biases within datasets, such as those used for training or testing. In addition, the individual case studies imply that co-creation and collaboration are more proactive while contribution and consultation are more reactive. The author draws this conclusion as co-creation and contribution generally involve technical and non-technical experts working together at the beginning of the ML system design process. Contribution and consultation, alternatively, are generally implemented once technical designers have already started or finished designing the ML system.[261]

Cross-Case Synthesis Conclusions

With the cross-case synthesis of the findings from the three individual case studies, the researcher can generally understand which unintended biases may appear in which JIP phase as well as which participatory design best practice can mitigate these unintended biases. For the first part of the cross-case synthesis, the study focuses on determining the overall unintended biases within the JIP phases. Nine of the eleven

[261] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 11-14.

analyzed unintended biases can generally appear within the second (collection) and third (processing and exploitation) phases of the JIP. Three of the eleven analyzed unintended biases can generally appear within the fifth (integration and dissemination) phase of the JIP. The study identifies these specific unintended biases in the fourth chapter as well as outlines them in the next section. Again, these conclusions are generalized from variables and respective sub-variables that the study identifies in the three collected case studies.

For the second part of the cross-case synthesis the study focuses on understanding which participatory design best practices can generally mitigate which unintended biases. Each case study aligns with one to two categories of participatory design. While the study identifies the presence of both co-creation and collaboration in the first case study, the researcher acknowledges that these are very similar strategies, and that collaboration can fall within co-creation.[262] As each case study generally aligns with a different participatory design best practice, the cross-referencing is relatively straightforward. The study determines that the unintended biases within each case study can be mitigated with the presence of the respective participatory design best practice. For the first case study, the co-creation participatory design best practice was used to try to mitigate unintended biases associated with the algorithm design.[263] For the second case study, the focus was on using contribution to try to mitigate unintended biases associated with datasets.[264] In the third case study, the use of consultation was used to try to influence companies to

[262] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 11.

[263] Lee, Kim, and Lizarondo, "A Human-Centered Approach to Algorithmic Services," 3365-3376.

[264] Sarin et al., "Crowdsource by Google," 1-3.

improve their ML systems' datasets as well as educate the public and inspire government regulatory action.[265] The study acknowledges that the use of only three case studies limits understanding of all the unintended biases these participatory design best practices could mitigate. These best practices may be able to mitigate additional unintended biases.

Research Question Conclusions

This study has one primary research question with three secondary research questions. The primary research question focuses on how the military intelligence community can help mitigate unintended biases in the JIP. The cross-case synthesis seeks to answer the primary research question. The secondary research questions focus on the following: 1) how ML influences situational awareness or decision-making; 2) what the indicators of unintended bias within ML systems; and 3) what the best practices are to mitigate unintended biases within ML systems. The literature review and the variable analysis within the case studies address the secondary research questions. This study addresses the answers to these research questions in the subsequent paragraphs.

For the first secondary research question, the study's literature review describes the current influence of ML systems within commercial industry's situational awareness and decision-making. It also analyzes each of the three case studies for JIP related tasks and unintended biases. The study suggests that having unintended bias in an ML system operating within the JIP can result in biased and inaccurate intelligence that negatively impacts situational understanding and decision-making. As per the case study variable analysis, the study identifies unintended biases predominately appearing in the second

[265] Buolamwini and Gebru, "Gender Shades," 1-15; Raji and Buolamwini, "Actionable Auditing," 1-7.

and third phases of the JIP. By having bias present in the collection phase as well as the processing and exploitation phase, this can cause detrimental impacts throughout the rest of the JIP as analysts use the collected and processed data for the analysis and production which they disseminate to decision-makers for feedback and evaluation.

Regarding the second secondary research question, the study identifies current literature describing indicators of unintended bias as well as identifies the presence of unintended biases within the individual case studies. The study specifically looks for the following unintended biases: sampling, measurement, label, negative set, framing effect, sample selection, confounding, design, sample treatment, human evaluation, and validation and test dataset.[266] In addition to identifying these types of unintended biases within ML systems, the study describes the indicators associated with these unintended biases in-depth in the literature review. Some of the notable indicators are reiterated in this section.

As outlined in Chander and Srinivasan's study, titled "Biases in AI Systems," indicators of unintended bias may appear throughout the various phases of the "AI/ML pipeline."[267] Regarding the "data-creation" phase, some indicators of unintended bias include the ML system being unable to perform the following: 1) apply data trends across all groups, 2) display data in different ways, and 3) properly or consistently label data, and 4) identify "negative instances."[268] In the "problem formulation" phase, an indicator

---

[266] Chander and Srinivasan, "Biases in AI Systems," 45-64.

[267] Ibid., 56.

[268] Chander and Srinivasan, "Biases in AI Systems," 48-50; Torralba and Efros, "Unbiased Look at Dataset Bias,"50; Mehrabi et al., "A Survey on Bias and Fairness in Machine Learning," 6.

of unintended bias may manifest as an ML system providing inconsistent query results.[269] In the next phase, "data analysis," an ML system may have indicators of unintended bias which causes results that "are not representative of the population intended to be analyzed," have an "omitted variable," or use a "proxy variable."[270] As for "design-related bias," the indicators are often difficult to identify and generally a technical designer will need to review them.[271] For the last phase, "evaluation/validation," indicators manifest as "inappropriate and disproportionate benchmarks" or showing only certain data or results to some people and not others.[272]

As for the last secondary research question, the study also describes current commercial industry best practices, namely participatory design strategies, to mitigate unintended biases as well as identifies the implementation of these strategies against unintended biases within the case studies. There is a multitude of technical mitigation strategies available, but the general military intelligence community will unlikely be able to execute these technical mitigation strategies; therefore, the study focuses on non-technical strategies. The study identifies the following participatory design best practices as ways the military intelligence community can help mitigate unintended biases within ML systems: "co-creation," "collaboration," "contribution," and "consultation."[273] These

---

[269] Chander and Srinivasan, "Biases in AI Systems," 51-52.

[270] Ibid., 52-53.

[271] Ibid., 53.

[272] Chander and Srinivasan, "Biases in AI Systems," 54; Mehrabi et al., "A Survey on Bias and Fairness in Machine Learning," 8.

[273] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 11.

best practices provide a framework for the military intelligence community to engage with technical designers, but while also using its respective subject matter expertise to improve ML systems used in the JIP.

It is not until the cross-case synthesis that the study is able to fully address the primary research question. The first part of the cross-case synthesis seeks to answer the latter part of the primary research question which is understanding which unintended biases may appear in which JIP phase. The synthesis reveals that most of the unintended biases appear in the second and third phases of the JIP. The unintended biases that appear in these two phases include sampling, measurement, label, negative set, framing effect, sample selection, design, human evaluation, and test dataset. Only confounding and sample treatment did not appear in these two phases. This assists those who are working with these ML systems to identify these unintended biases within the second and third phases of the JIP.

The second part of the cross-case synthesis looks to answer the first part of the primary research question that focuses on how the military intelligence community can mitigate these unintended biases in its ML systems. As per the cross-case synthesis, there are different best practices which one may be able to employ depending on the situation. If the ML system is within its initial stages of development, then one may implement co-creation or contribution; however, if the ML system is already in operation, then contribution or consultation might be more appropriate.[274] Ultimately, the study concludes that the military intelligence community can help mitigate unintended biases

[274] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 11.

within ML systems by not only understanding the types and indicators of unintended biases, but also implementing the different participatory design strategies to help improve the ML systems operating in the JIP. This study provides this community an awareness and initial framework for identifying unintended biases, predominately within the second and third phases of the JIP and implementing participatory design mitigation strategies; depending upon where an ML system is in its design process and how much influence the users have in the design.

Recommendations

The study has several recommendations for improving the study, conducting follow-on research, and implementing the findings within the military. The section below discusses each of these recommendations. The researcher discusses them in the following order: 1) recommendations to improve the study; 2) recommendations for future research; and 3) recommendations for implementation within the military.

Recommendations for Study Improvement

The study acknowledges that this is predominately an exploratory study due to the limitation and delimitations specified in the first chapter. First and foremost, the researcher recommends gaining access to propriety and/or classified information to improve the study's findings. To identify which unintended biases appear exactly in which phases of the JIP, the study would likely need access to proprietary and classified information. Secondly, the researcher recommends diversifying and increasing the number of case studies. For example, the first two case studies included Carnegie Mellon researchers, which could influence the findings of this study. Lastly, the researcher

recommends either extending the study's research period or having an individual with greater ML technical expertise repeat the study to allow for greater data collection and more detailed data analysis.

Recommendations for Future Research

The study reveals various opportunities for future research. First, the researcher recommends measuring the effectiveness of the participatory design best practices through experimentation by comparing the results of when designers do and do not use it. In addition, the researcher recommends analyzing the effectiveness of each participatory design on individual unintended biases. Secondly, the researcher recommends studying additional best practices, either technical or non-technical, that commercial industry has used to mitigate unintended biases. Thirdly, the researcher recommends also analyzing best practices adopted by the public sector to mitigate unintended biases. Lastly, the researcher recommends analyzing how the military currently designs its ML systems to look for possible improvements.

Recommendations for Implementation in the Military

The study recommends the military intelligence community adopt these participatory design best practices to help mitigate unintended biases within their ML systems. Ideally, the military intelligence community will co-create or collaborate with technical designers on future ML systems for the JIP process; however, if the ML systems are already in design or implemented, then the study suggests the military intelligence community use contribution and consultation best practices. As two of the three case studies center on mitigating bias within datasets, it is recommended that the

community focus efforts similarly by improving awareness of unintended biases in training and testing data as well as the respective participatory design mitigation strategies to improve ML system performance. Also, the community should not forget about the unintended biases which may manifest from an algorithm; however, unless the community is directly involved in co-creating or collaborating with the technical designers on the algorithm, this type of unintended bias is difficult for the community to detect and provide feedback to the technical designers.[275]

Understanding that the DOD has already suggested that it will focus on a human-centered approach, this study helps further this endeavor by educating the general military intelligence community as well as exposing this community to various forms of participatory design and how commercial industry implements them. As for the *DOD AI Education Strategy*, the strategy does not go into depth as to what exactly will be covered within the curriculum.[276] As the DOD executes its *DOD AI Education Strategy*, the researcher recommends the military covers the basic components of this study's material (e.g., unintended bias types and indicators, participatory design strategies, ML influence on situational understanding or decision-making) within this "required instruction."[277]

---

[275] Berditchevskaia, Malliaraki, and Peach, "Participatory AI for Humanitarian Innovation," 11.

[276] JAIC, "2020 Department of Defense Artificial Intelligence Education Strategy," ii.

[277] Ibid.

Summary

This chapter covers the conclusions and recommendations of this study. The study provides an overview of the conclusions from the case study variable analysis and the cross-case synthesis. In addition, it uses this analysis and synthesis to answer the primary and secondary research questions. After providing the conclusions, the study supplies a myriad of recommendations for study improvement, future research topics, and ways to implement the study's findings into the military intelligence community. Overall, the purpose of this study is to support the DOD with implementing its ethical principal of equitability in ML systems as well as to help the military intelligence community understand how it can contribute to the improvement of these systems for the ultimate goal of providing equitable intelligence in a more expedient manner.[278] In summary, the study recommends the military intelligence community adopt participatory design strategies to help mitigate unintended biases which may appear in ML systems operating throughout the various stages of the JIP. With the military intelligence community's adoption of these mitigation strategies as best practices, this will likely help improve these ML systems over time and provide more equitable intelligence to the greater military community.

[278] DOD, "DOD Adopts Ethical Principles for Artificial Intelligence."

BIBLIOGRAPHY

412 Food Rescue. “412 Food Rescue.” Accessed April 11, 2022. https://412foodrescue.org/.

———. “Food Rescue Hero.” Accessed April 11, 2022. https://412foodrescue.org/programs/foodrescuehero/.

———. “What We Do.” Accessed April 11, 2022. https://412foodrescue.org/about-us/what-we-do/.

AI.Business “Can Artificial Intelligence Reduce Food Waste?” AI. January 25, 2017. https://ai.business/2017/01/25/can-artificial-intelligence-reduce-food-waste/.

Allen, Greg. “Understanding Artificial Intelligence Technology.” Joint Artificial Intelligence Center. April 2020. https://www.ai.mil/docs/Understanding%20AI%20Technology.pdf.

Allen, John R., and Darrell M. West. “How Artificial Intelligence is Transforming the World.” The Brookings Institution. 2018. https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/.

Amazon. “Detecting and Analyzing Faces.” AWS. Accessed April 15, 2022. https://docs.aws.amazon.com/rekognition/latest/dg/faces.html.

Apple. “Food Rescue Hero.” App Store Preview. Accessed April 11, 2022. https://apps.apple.com/us/app/food-rescue-hero/id1518660483.

Batra, Anurag. “How to Help Make AI Systems More International and Inclusive with Google’s Crowdsource App.” Accelerate with Google. Accessed April 11, 2022. https://accelerate.withgoogle.com/stories/how-to-help-make-AI-systems-more-international-and-inclusive-with-googles-crowdsource-app.

Batra, Anurag, and Parker Barnes. “Adding Diversity to Images with Open Images Extended.” *Google AI Blog* (blog). *Google Research*, December 7, 2018. https://ai.googleblog.com/2018/12/adding-diversity-to-images-with-open.html.

BBC. “Google Apologises for Photo App’s Racist Blunder.” *BBC News*, July 1, 2015. https://www.bbc.com/news/technology-33347866.

Ben-Yair, Shimrit. “Your Photos, Your Memories, Your Way.” *The Keyword* (blog). *Google*, May 18, 2021. https://blog.google/products/photos/new-memories-features-look-back/.

Berditchevskaia, Aleks, Eirini Malliaraki, and Kathy Peach. *Participatory AI for Humanitarian Innovation: A Briefing Paper.* London, UK: Nesta, 2021. https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf.

Blum, Avrim. "Machine Learning Theory." Essay, Carnegie Melon University, School of Computer Science, 2007. http://www.cs.cmu.edu/afs/cs/user/avrim/www/Talks/mlt.pdf.

Brainard, Lael. "What Are We Learning about Artificial Intelligence in Financial Services?" Speech, Fintech and the New Financial Landscape, Philadelphia, PA, November 13, 2018. https://www.federalreserve.gov/newsevents/speech/brainard20181113a.htm.

Buolamwini, Joy. "When the Robot Doesn't See Dark Skin." *The New York Times,* June 21, 2018. https://www.nytimes.com/2018/06/21/opinion/facial-analysis-technology-bias.html.

Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of *Machine Learning Research* 81 (2018): 1-15. Conference on Fairness, Accountability, and Transparency, New York University, New York City, February 23-24, 2018. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

Cambridge Dictionary Editors. "Acquisition." Cambridge Dictionary, Cambridge University Press. Accessed April 22, 2022. https://dictionary.cambridge.org/us/dictionary/english/acquisition.

———. "Stakeholder." Cambridge Dictionary, Cambridge University Press. Accessed April 22, 2022. https://dictionary.cambridge.org/us/dictionary/english/stakeholder.

CGSC Learning Resource Center. Combined Arms Research Library. E-mail submission. May 6, 2022. Reviewed for grammar, punctuation, and clarity of expression.

Chairman of the Joint Chiefs of Staff. Joint Publication 2-0, *Joint Intelligence*. Washington, DC: Joint Chiefs of Staff, October 2013. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf.

Chander, Ajay, and Ramya Srinivasan. "Biases in AI Systems." *ACM Queue* 19, no. 2 (2021): 45-64. https://dl.acm.org/doi/pdf/10.1145/3466132.3466134.

Dastin, Jeffrey, and Munsif Vengattil. "Microsoft Bans Face-Recognition Sales to Police as Big Tech Reacts to Protests." *Reuters,* June 11, 2020. https://www.reuters.com/article/us-microsoft-facial-recognition/microsoft-bans-face-recognition-sales-to-police-as-big-tech-reacts-to-protests-idUSKBN23I2T6.

Defense Innovation Board. “AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense.” Supporting Document. U.S. Department of Defense, Washington, DC, October 2019. https://media.defense.gov/2019/Oct/31/2002204459/-1/-1/0/DIB_AI_PRINCIPLES_SUPPORTING_DOCUMENT.PDF.

Department of Defense. “DOD Adopts Ethical Principles for Artificial Intelligence.” February 24, 2020. https://www.defense.gov/News/Releases/Release/Article/2091996/DOD-adopts-ethical-principles-for-artificial-intelligence/.

———. *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*. Washington, DC: Department of Defense, 2018. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

———. “U.S. Department of Defense.” Accessed October 12, 2021. https://www.defense.gov/.

Doshi, Tulsee. “Introducing the Inclusive Images Competition.” *Google AI Blog* (blog). *Google Research*, September 6, 2018. https://ai.googleblog.com/2018/09/introducing-inclusive-images-competition.html.

Dougherty, Conor. “Google Photos Mistakenly Labels Black People ‘Gorillas’.” *Business, Innovation, Technology, Society* (blog). *The New York Times*, July 1, 2015. https://bits.blogs.nytimes.com/2015/07/01/google-photos-mistakenly-labels-black-people-gorillas/.

Face++. “Detect API.” Doc Center. Accessed April 15, 2022. https://console.faceplusplus.com/documents/5679127.

Fletcher, Richard, Audace Nakeshimana, and Olusubomi Olubeko. “Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health.” *Frontiers in Artificial Intelligence* 3 (April 2021): 1-17. https://doi.org/10.3389/frai.2020.561802.

Gasson, Susan. “Human-Centered vs. User-Centered Approaches to Information System Design.” *The Journal of Information Technology Theory and Application* (JITTA) 5, no. 2 (2003): 29-46.

Gender Shades. “IBM Response to ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification’.” MIT Media Lab. January 2018. http://gendershades.org/docs/ibm.pdf.

Google. “412 Food Rescue.” Google Play. Accessed April 11, 2022. https://play.google.com/store/apps/details?id=com.fouronetwo.foodrescue&referrer=utm_source%3Dinstagramadsummer18%26utm_medium%3DSocial.

———. “Crowdsource.” Google Play. Accessed April 11, 2022.https://play.google.com/store/apps/details?id=com.google.android.apps.village.boond.

———. “Food Rescue Hero.” Google Play. Accessed April 11, 2022. https://play.google.com/store/apps/details?id=org.foodrescuehero.app.

———. “How Crowdsource Works.” Crowdsource by Google. Accessed April 13, 2022. https://crowdsource.google.com/about/how-it-works/.

———. “Open Images Extended - Crowdsourced.” Google LLC. Accessed April 11, 2022. https://research.google/static/documents/datasets/open-images-extended-crowdsourced.pdf.

———. “Open Images Extended - Crowdsourced.” Google Research. 2018. Accessed April 11, 2022. https://research.google/tools/datasets/open-images-extended-crowdsourced/.

———. “Overview of Open Images V6.” Open Images Dataset V6. Accessed April 11, 2022. https://storage.googleapis.com/openimages/web/factsfigures.html.

———. “The Home for Your Memories.” Google Photos. Accessed April 13, 2022. https://www.google.com/photos/about/.

Greene, Jay. “Microsoft Won’t Sell Police Its Facial-Recognition Technology, Following Similar Moves by Amazon and IBM.” *The Washington Post,* June 11, 2020. https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/.

Guynn, Jessica. “Google Photos Labeled Black People ‘Gorillas’.” *USA Today*, July 1, 2015. https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/.

Hardesty, Larry. “Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems.” *MIT News,* February 11, 2018. https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212.

Harwell, Drew. “Federal Study Confirms Racial Bias of Many Facial-Recognition Systems, Cast Doubt on Their Expanding Use.” *The Washington Post,* December 29, 2019. https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/.

Heuer, Jr., Richards J. *Psychology of Intelligence.* Langley, VA: Central Intelligence Agency, Center for the Study of Intelligence, 1999. https://www.ialeia.org/docs/Psychology_of_Intelligence_Analysis.pdf.

IBM. “AI Fairness 360.” Accessed November 16, 2021. https://aif360.mybluemix.net/.

———. “Build Smarter Supply Chains with AI and Blockchain.” Accessed May 1, 2022. https://www.ibm.com/supply-chain.

———. “Building Trust in AI.” Accessed February 11, 2022. https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/building-trust-in-ai.html.

———. “IBM CEO’s Letter to Congress on Racial Justice Reform.” *THINKPolicy Blog* (blog), June 8, 2020. https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/?mhsrc=ibmsearch_a&mhq=congress%20facial%20recognition.

IBM Research Editorial Staff. “IBM to Release World’s Largest Annotation Dataset for Studying Bias in Facial Analysis.” *IBM Research Blog* (blog). *IBM*, June 27, 2018. https://www.ibm.com/blogs/research/2018/06/ai-facial-analytics/.

International Organization for Standardization (ISO). “Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems.” Online Browsing Platform. 2019. https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:en.

Johnson, Khari. “Congress Moves toward Facial Recognition Regulation.” Venture Beat. January 15, 2020. https://venturebeat.com/2020/01/15/congress-moves-toward-facial-recognition-regulation/.

———. “IBM Walked Away from Facial Recognition. What about Amazon and Microsoft?” Venture Beat. June 10, 2020. https://venturebeat.com/2020/06/10/ibm-walked-away-from-facial-recognition-what-about-amazon-and-microsoft/.

Joint Artificial Intelligence Center. “Department of Defense Joint Artificial Intelligence Center Responsible AI Champions Pilot.” 2020. https://www.ai.mil/docs/08_21_20_responsible_ai_champions_pilot.pdf.

———. “Joint Artificial Intelligence Center.” Accessed November 21, 2021. https://www.ai.mil/.

Joint Artificial Intelligence Center DoD Chief Information Officer. “2020 Department of Defense Artificial Intelligence Education Strategy.” Joint Artificial Intelligence Center, Washington, DC, September 2020. https://www.ai.mil/docs/2020_DoD_AI_Training_and_Education_Strategy_and_Infographic_10_27_20.pdf.

Kairos. “API Reference.” Developing with Kairos. Accessed April 15, 2022. https://www.kairos.com/docs/api/.

Kelly, Kate. “How One App Saved Over 40 Million Pounds of Food from the Landfill.” *Tech Soup* (blog). *Tech Soup*, March 15, 2021. https://blog.techsoup.org/posts/how-one-app-saved-over-40-million-pounds-of-food-from-the-landfill.

Learned-Miller, Erik, Vicente Ordonez, Jamie Morgenstern, and Joy Buolamwini. *Facial Recognition Technologies in the Wild: A Call for a Federal Office*. Cambridge, MA: Algorithmic Justice League, May 29, 2020. https://assets.website-files.com/5e027ca188c99e3515b404b7/5ed1145952bc185203f3d009_FRTsFederalOfficeMay2020.pdf.

Lee, Kyung Min, Ji Tae Kim, and Leah Lizarondo. “A Human-Centered Approach to Algorithmic Services: Considerations for Fair and Motivating Smart Community Service Management that Allocates Donations to Non-Profit Organizations.” Paper presented at the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, May 6-11, 2017. https://dl.acm.org/doi/pdf/10.1145/3025453.3025884.

Lohr, Steve. “Facial Recognition is Accurate, if You’re a White Guy.” *The New York Times*, February 9, 2018. https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html?action=click&module=RelatedLinks&pgtype=Article.

Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. “A Survey on Bias and Fairness in Machine Learning.” Arxiv. Cornell University, 2022. https://arxiv.org/pdf/1908.09635.pdf.

MEGVII. “Face++.” Accessed April 15, 2022. https://www.faceplusplus.com/.

———. “Notice: Newer Version of Face Detect API.” *Face++* (blog), June 27, 2018. https://www.faceplusplus.com/blog/article/newer-version-face-detect-api/?cate=product_news.

Merriam-Webster Dictionary Editors. “Best Practice.” Merriam-Webster Dictionary, Merriam-Webster Incorporated. Accessed January 25, 2021. https://www.merriam-webster.com/dictionary/best%20practice.

Microsoft. “Face API.” Microsoft Azure. Accessed April 15, 2022. https://azure.microsoft.com/en-us/services/cognitive-services/face/.

———. “What is Azure Face Service?” March 2, 2022. https://docs.microsoft.com/en-us/azure/cognitive-services/face/overview.

MIT Media Lab. “Gender Shades.” 2018. http://gendershades.org/overview.html.

Office of the Director of National Intelligence. “Artificial Intelligence Ethics Framework for the Intelligence Community.” Version 1. Director of National Intelligence, Washington, DC, June 2020.

https://www.dni.gov/files/ODNI/documents/AI_Ethics_Framework_for_the_Intelligence_Community_10.pdf.

———. "Members of the IC." Accessed October 12, 2021. https://www.dni.gov/index.php/what-we-do/members-of-the-ic.

———. *The AIM Initiative: A Strategy for Augmenting Intelligence Using Machines*. Washington, DC: Director of National Intelligence, 2019. https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf.

Osoba, Osonde A., and William Wesler, IV. *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. Santa Monica, CA: RAND Corporation, 2017. https://www.rand.org/pubs/research_reports/RR1744.html.

Ovide, Shira. "A Case for Banning Facial Recognition." *The New York Times,* updated August 1, 2021. https://www.nytimes.com/2020/06/09/technology/facial-recognition-software.html.

Puri, Ruchir. "Mitigating Bias in AI Models." *IBM Research Blog* (blog). *IBM*, February 6, 2018. https://www.ibm.com/blogs/research/2018/02/mitigating-bias-ai-models/.

Raji, Inioluwa Deborah, and Joy Buolamwini. "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products." *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI/ACM) (2019): 429-435.* Conference on AI, Ethics, and Society, Honolulu, Hawaii, January 27-28, 2019. https://www.thetalkingmachines.com/sites/default/files/2019-02/aies-19_paper_223.pdf.

Roach, John. "Microsoft Improves Facial Recognition Technology to Perform Well across All Skin Tones, Genders." *The AI Blog* (blog). *Microsoft*, June 26, 2018. https://blogs.microsoft.com/ai/gender-skin-tone-facial-recognition-improvement/.

Sabhrwal, Anil. "Picture This: A Fresh Approach to Photos." *The Keyword* (blog). *Google*, May 28, 2015. https://blog.google/products/photos/picture-this-fresh-approach-to-photos/.

Sanchez, Galen. Proofread (Microsoft Word, CGSC, Ft Leavenworth, November 28, 2021; November 30, 2021; March 24, 2022; March 31, 2022; April 17).

Sarin, Supheakmungkol, Knot Pipatsrisawat, Khiem Pham, Anurag Batra, and Lu´ıs Valente. "Crowdsource by Google: A Platform for Collecting Inclusive and Representative Machine Learning Data." Paper presented in the Work in Progress and Demo Track at the seventh AAAI Conference on Human Computation and Crowdsourcing 2019, Skamania Lodge, WA, October 28-30, 2019. https://storage.googleapis.com/pub-tools-public-publication-data/pdf/55f12b7d77b32432b36970ac4e9ee57cb546ca7f.pdf.

Schupak, Amanda. “Google Apologizes for Mis-tagging Photos of African Americans.” *CBS News*, July 1, 2015. https://www.cbsnews.com/news/google-photos-labeled-pics-of-african-americans-as-gorillas/.

Secretary of Defense. *Summary of the 2018 National Defense Strategy.* Washington, DC: Department of Defense, 2018. https://DOD.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf.

Sexton, Jason. Proofread (Microsoft Word, CGSC, Ft Leavenworth, October 12, 2021).

Simonite, Tom. “Google Turns to Users to Improve its AI Chops Outside the US.” *Wired*, April 5, 2018. https://www.wired.com/story/google-turns-to-users-to-improve-its-ai-chops-outside-the-us/.

Singer, Natasha. “Amazon is Pushing Facial Technology That a Study Says Could be Biased.” *The New York Times,* January 24, 2019. https://www.nytimes.com/2019/01/24/technology/amazon-facial-technology-study.html.

Smith, Brad. “Finally, Progress on Regulating Facial Recognition.” *Microsoft on the Issues* (blog). *Microsoft*, March 31, 2020. https://blogs.microsoft.com/on-the-issues/2020/03/31/washington-facial-recognition-legislation/.

Swanson, E. Justin. “Upload the Pictures, and Let Google Photos Do the Rest.” *The New York Times*, June 3, 2015. https://www.nytimes.com/2015/06/04/technology/personaltech/upload-the-pictures-and-let-google-photos-do-the-rest.html.

Team Crowdsource, “Welcome to the Crowdsource by Google Blog!” *Crowdsource by Google* (blog). *Google*, May 2020. https://crowdsource.google.com/about/blog/welcome-to-the-crowdsource-blog/.

Torralba, Antonio and Alexei Efros. “Unbiased Look at Dataset Bias.” In the *IEEE Conference* Proceedings on *Computer Vision and Pattern Recognition*, IEEE, June 2011: 1521-1528; https://ieeexplore.ieee.org/document/5995347. Quoted in Ajay Chander and Ramya Srinivasan, “Biases in AI Systems,” *ACM Queue* 19, no. 2 (2021): 50, https://dl.acm.org/doi/pdf/10.1145/3466132.3466134.

Turabian, Kate L. *A Manual for Writers of Research Papers, Theses, and Dissertations*. 9th ed. Chicago: The University of Chicago Press, 2018.

US Congress. House Committee on Oversight and Reform. *Facial Recognition Technology (Part 1): Its Impact on our Civil Rights and Liberties*. 116th Cong., May 22, 2019. https://oversight.house.gov/legislation/hearings/facial-recognition-technology-part-1-its-impact-on-our-civil-rights-and.

———. *Facial Recognition Technology (Part 1): Its Impact on Our Civil Rights and Liberties, Written Testimony of Joy Buolamwini*, 116th Cong., May 22, 2019, https://docs.house.gov/meetings/GO/GO00/20190522/109521/HHRG-116-GO00-Wstate-BuolamwiniJ-20190522.pdf.

US Congress, Senate. Facial Recognition and Biometric Technology Moratorium Act of 2021. S. Res. 117th Cong., 1st sess. (June 15, 2021): S.2052. https://www.congress.gov/bill/117th-congress/senate-bill/2052/all-info?r=1&s=1.

U.S. President. *National Security Strategy of the United States of America*. Washington, DC: The White House, December 2017. https://history.defense.gov/Portals/70/Documents/nss/NSS2017.pdf?ver=CnFwURrw09pJ0q5EogFpwg%3d%3d.

Vaughan, Jennifer. "Making Better Use of the Crowd: How Crowdsourcing can Advance Machine Learning Research." *Journal of Machine Learning Research* 18 (2018): 1-46. https://www.jmlr.org/papers/volume18/17-234/17-234.pdf.

Vincent, James. "IBM Hopes to Fight Bias in Facial Recognition with New Diverse Dataset." *The Verge,* June 27, 2018. https://www.theverge.com/2018/6/27/17509400/facial-recognition-bias-ibm-data-training.

Wiggers, Kyle. "Google's Inclusive Images Competition Spurs Development of Less Biased Image Classification AI." Venture Beat. December 2, 2018. https://venturebeat.com/2018/12/02/googles-inclusive-images-competition-spurs-development-of-less-biased-image-classification-ai/.

Wood, Matt. "Thoughts on Recent Research Paper and Associated Article on Amazon Rekognition." *AWS Machine Learning Blog* (blog). *AWS*, January 26, 2019, https://aws.amazon.com/blogs/machine-learningf/thoughts-on-recent-research-paper-and-associated-article-on-amazon-rekognition/.

Yin, Robert. *Case Study Research and Applications.* Thousand Oaks, CA: SAGE Publications, Inc. 2018.