Transforming in Contact:

Practical Ethics, Accountability, and Governance for Artificial Intelligence, Autonomy, and Information Work Across Government

by Bensson Samuel

Crises today develop quickly, demanding that governments rapidly sense, interpret, and make decisions under high pressure. This paper offers a targeted framework to balance speed with accountability, describing practical steps for interagency cooperation: shared data governance, model validation, lifecycle certification, auditable logs, pre-approved information playbooks, a joint information cell, ongoing ethical oversight, clear delegation, rules for private sector involvement, a central model registry, and a phased pilot plan. Although AI, autonomous systems, and information tools can speed up joint responses, they pose significant ethical, legal, and accountability issues for both civilian and military sectors. Without common technical standards and audit rules, agencies risk deadlock in joint crises. By establishing shared standards, ensuring auditability, and maintaining strong ethical guidelines, interagency operations can be both swift and responsible.

Ukraine's operational adaptations since 2022 demonstrate how decentralized, user-driven innovation, such as fielding commercial drones, mesh networking, small-unit sensors, and integrated open-source intelligence, can produce effects that surpass those of any single technology.[1] This underscores the importance of governance mechanisms that can effectively guide the integration and deployment of such innovations while ensuring accountability and ethical standards. The United States (U.S.) Army's Transforming in Contact concept also emphasizes decentralized decision-making, rapid learning cycles, and formations that sustain operations under contested and degraded communications.[2] In the current-day crisis, there is a need for governance mechanisms that ensure accountability and ethical guidelines while maintaining the flexibility of decentralized decision-

Dr. Bensson Samuel is a board-certified physician in internal medicine with specialized experience in critical care. He holds a Master of Laws in international and commercial law and a Master of Public Health, and he has completed United Nations Executive Diploma training in diplomatic practice and human rights. He also holds an Oxford Executive Leadership certificate from Saïd Business School and a Harvard health care leadership certificate, adding focused expertise in strategy, economics, and organizational change in health systems. His work examines how economic forces, legal frameworks, and global governance shape clinical decision-making, access to care, and resource allocation. His academic interests include health economics, global health diplomacy, and the intersection of law, policy, and frontline clinical practice.

making.

Modern crises condense sensing, interpretation, and action into shorter timeframes than in previous eras, benefiting organizations capable of learning on the fly.[3] Complex disaster responses illustrate the coordination challenges that arise when multiple agencies use different analytic tools and imperfect information sharing to make overlapping decisions that are ineffective.[4] Agencies across diplomacy, defense, intelligence, homeland security, justice, and development must act together quickly, and they increasingly rely on AI, autonomous systems, and information tools to support those actions.[5] At the same time, these technologies can obscure decision pathways and decision flowcharts, diffuse responsibility across public and private actors, and raise novel legal and ethical questions that governance must address.[6] With new technologies, previously established response protocols have been formulated without taking into account the indecisiveness that can occur in the response process.

These capabilities raise strategic, operational, and ethical questions that require careful governance. Any framework for their use must balance mission effectiveness with accountability, legality, and the preservation of human decision-making authority. Good governance for these capabilities should preserve civilian control, protect professional judgment, and ensure legal compliance while enabling prudent, time-sensitive risk-taking in contested environments.[7] To understand how emerging technologies are reshaping interagency governance, responsibility, and operational coordination, one must examine the smaller components involved in this process. To understand the roles, one has to consider AI-enabled decision support, autonomous systems, information operations, and private-sector technologies can improve speed, situational awareness, and mission effectiveness; nevertheless, they also create serious risks involving bias, legal responsibility, safety, public trust, and oversight.

Across these areas, the central challenge is how agencies can use advanced technologies quickly and effectively while still preserving accountability, human judgment, ethical standards, and lawful decision-making. The discussion therefore connects technical governance, moral responsibility, legal oversight, public-private cooperation, allied standards, measurable performance, and risk management into one broader argument: effective interagency technology governance requires clear authority, validated systems, human accountability, transparent audit trails, continuous review, and practical safeguards that allow innovation without sacrificing democratic control or public trust.

To understand how emerging technologies are reshaping interagency governance, responsibility, and operational coordination, one must examine the smaller components involved in this process.

AI-Enabled Decision Support: Balancing Speed, Accuracy, and Responsibility

AI-powered decision-support systems can pull together data from many sources, find patterns, and offer options faster than human teams alone.[8] In interagency work, these systems can help prioritize aid, predict migration and refugee flows, assess risks to diplomatic missions, and connect cyber-attribution data from different sources.[9] When models and dashboards are well-designed and validated, they help agencies share a common view and cut down on coordination problems.[10] AI can also speed up triage in big, fast-moving crises by flagging urgent information for people to

review and helping allocate limited resources.[11] A crucial component of these systems is the "human-machine feedback cycle." In this cycle, AI suggestions are promptly evaluated by human analysts who interpret the outputs and provide necessary adjustments based on real-time situational awareness. This ensures decisions are not only rapid but also informed, helping maintain the accountability and reliability of actions taken. Standardized outputs make decisions more consistent across agencies and regions, especially when governance and documentation clearly show model limits.[12]

When several agencies use algorithmic recommendations, it can be unclear who is legally or politically responsible for actions influenced by automated outputs.

However, these systems also pose real risks for coordination and accountability between agencies. If models are trained on old or biased data, they may fail when situations change or if someone tries to manipulate the inputs, leading to unreliable results that can be dangerous without proper oversight.[13] Training data that excludes certain groups or situations, such as refugees from specific regions, can lead to unfair outcomes and worsen injustice when those results guide resource allocation.[14, 15] When several agencies use algorithmic recommendations, it can be unclear who is legally or politically responsible for actions influenced by automated outputs.[16] Finally, adversaries could attempt to corrupt or forge shared data streams if agencies lack strict rules for data origin, integrity, and defense against adversarial attacks.[17, 18]

To handle these risks without slowing down, agencies need to put clear governance steps in place.

First, establish a required interagency data governance framework with rules for data provenance, tagging, minimum subgroup sample sizes, and required metadata. Every model shared between agencies should include a data dictionary, evidence of where the training data came from, subgroup performance metrics, and a statement of its limitations (a model card).[19] For instance, if a model predicting refugee surges is based on outdated data, it could misallocate resources and exacerbate a crisis. With proper governance, by updating data sources and ensuring rigorous testing, this failure could be averted, demonstrating clear benefits.

Second, make sure models are tested in real scenarios, checked for adversarial attacks, and monitored regularly. Validation should include tests for data changes, stress tests for attacks, and regular peer reviews by both technical and subject-matter experts.[20, 21]

Third, establish a clear policy outlining when human involvement is required in the decision-making process. This policy should align the level of automation with the risk and make it clear whether outputs are advice, recommendations, or direct instructions. For important decisions, a specific person should sign off and take legal and policy.[22, 23]

Fourth, require audit trails that cannot be changed or tampered with, recording all inputs, model versions, outputs, and key human decisions. These logs should be kept in a way that authorized oversight groups can review them after the fact.[24]

Finally, back up these technical steps with training and clear procedures so operators know the models' limits and when to bring in a human decision-maker.[25, 26]

Autonomous Systems: Safety, Responsibility, and Shared Spaces

Autonomous platforms are increasingly used for inspection, monitoring, surveillance, transport, and other tasks across civilian and military missions.[27] Civilian agencies have

expanded autonomy for border monitoring, disaster assessment, and infrastructure inspection, which raises the likelihood that government and commercial autonomous systems will operate in proximity to military systems, first responders, and crowded urban environments.[28] When different organizations acquire autonomy under divergent safety and assurance regimes, inconsistent testing and certification expectations can create operational risk at the boundaries between systems.[29] For example, equipment procured as commercial off-the-shelf (COTS) by a civilian agency may not undergo the same software assurance and lifecycle testing that DoD programs typically require, producing asymmetries in traceability and reliability.[30, 31] These procurement and assurance gaps complicate the question of legal responsibility when autonomous systems cause harm, whether a collision, a damaging false positive, or another failure, because accountability can be distributed across acquirers, operators, contractors, and vendors.[32] To clarify these responsibility hand-offs, a detailed matrix should be developed outlining the transition points where legal accountability shifts between entities. This matrix would describe, for example, points where responsibility passes from model developers to data providers, from operators to oversight agencies, and from government agencies to private vendors. Such a tool would serve to ensure all parties involved understand their obligations and can respond effectively in the event of an incident.

Shared operating spaces require interoperable safety procedures and real-time deconfliction mechanisms that span agencies and sectors.[33] Cities and complex disaster zones are particularly challenging because of dynamic obstacles, mixed human and machine traffic, and the need for predictable fail-safe behaviors from autonomous assets.[34] To reduce ambiguity and improve safety, agencies should create an Interagency Autonomy Standards Board that defines autonomy modes, minimum human-in-the-loop requirements for given mission sets, and cross-cutting safety certification benchmarks.[35] The board would bring together technical experts, program managers, legal counsel, privacy and civil liberties officers, operational stakeholders, social scientists, and community representatives to ensure standards are technically rigorous and socially legitimate.[36] Additionally, incorporating input from civil society and public consultations can enhance the board's legitimacy and ensure that public concerns about the use of shared urban airspace are addressed proactively. Engaging with communities will help build trust, provide transparency, and preempt any potential trust gaps related to the deployment and operation of autonomous systems in urban settings.

Shared operating spaces require interoperable safety procedures and real-time deconfliction mechanisms that span agencies and sectors.

Additionally, agencies should mplement a lifecycle certification strategy that encompasses procurement, deployment, maintenance, and software updates. They should also ensure that software modifications prompt necessary revalidation or recertification as applicable. Certification records and provenance data should be preserved to enable independent thirdparty audits, and contracts should include clauses that provide investigators with access to vendor artifacts under prearranged authorities.[37, 38] Agencies should also agree on shared incident investigation protocols—standardized evidencepreservation practices for sensor logs, telemetry, and software snapshots—and on independent investigative panels with the technical capacity to reconstruct system behavior

after an event.[39] Favoring modular autonomy stacks that separate sensing, planning, and actuation makes forensic reconstruction simpler and supports explainability, and investing in explainable AI techniques for perception components aids postincident analysis.[40, 41]

Information operations now sit at the intersection of public diplomacy, national security, and domestic information ecosystems.

Information operations now sit at the intersection of public diplomacy, national security, and domestic information ecosystems. Open-source intelligence, social amplification, synthetic media, and targeted narratives can have a strategic effect far beyond their tactical origins.[42] For domestic audiences, U.S. law restricts government efforts to influence and requires agencies to maintain clear separations between activities aimed at foreign audiences and communications that could reach or affect domestic populations.[43, 44] Mistaken attribution, premature public claims, or perceived manipulation in this context can result in significant domestic legal implications and damage to public trust, making careful standards for domestic communication and operations essential. For foreign audiences, the potential for diplomatic blowback is heightened by errors in attribution, premature public claims, or perceived manipulation in information activities. Such mistakes can damage diplomatic relationships, making careful standards for attribution and release crucial.[45]

To act quickly and lawfully when information crises emerge, agencies should prepare pre-authorized playbooks for predictable contingencies such as foreign disinformation campaigns, disaster-related misinformation, and embassy cyber incidents that include pre-vetted messaging templates, attribution standards, assigned roles, and legal checklists.[46, 47] Establishing a standing joint information cell that brings together State, the Department of War (DOW), DHS, USAID, and intelligence representatives can support legal review, message coherence, and rapid deconfliction in high-tempo situations.[48] Such a cell should adopt interagency confirmation protocols that assign confidence levels to open source intelligence (OSINT), require documented source provenance, and set thresholds for public release so that decision-makers can weigh risk and credibility under time pressure.[49, 50] Where legally permissible, publishing post-incident analyses that explain methods, limitations, and uncertainty helps build public resilience and reduces the adversary's ability to exploit ambiguity.[51] Each playbook should be accompanied by a quantitative indicator, such as achieving an average attribution confidence score within two hours, to ensure that guidance becomes a trackable commitment.

Moral Governance: Putting Values into Practice

Moral governance must be practical and forwardlooking, shaping system design, procurement, training, and operations rather than remaining an afterthought.[52] One practical measure is a standing interagency "ethics horizon" board composed of ethicists, legal experts, technologists, and practitioners to identify emerging risks, provide rapid advisory opinions in crises, and keep a public register of guidance for program offices and operators.[53] Agencies should require ethics impact assessments for new capabilities—akin to privacy impact assessments—that identify foreseeable harms, mitigation steps, monitoring metrics, and thresholds for suspension or rollback.[54, 55] Embedding dedicated ethics and legal advisors inside program management offices and operational units helps teams

get realtime consultation and produce predeployment attestations of legal and ethical review.[56] Training should incorporate "moral triage" exercises and difficult ethical scenarios—especially those that replicate degraded communications and compressed timelines—so decisionmakers practice tradeoffs and learn to use formal authorization processes under stress.[57, 58] To measure progress, adopt simple, trackable metrics such as the number of ethically escalated decisions in exercises or the proportion of deployments covered by a completed ethics impact assessment, which helps make the value of ethics resources visible to budget holders.[59] Finally, ethical review processes should be iterative and timebounded: rapid advisory inputs during crisis operations should be coupled with postdeployment reviews that feed lessons back into procurement, certification, and training cycles.[60, 61]

Accountability: Mapping Authority and Enabling Oversight

Accountability starts with clarity about who may act, when they may act, and what evidence will exist afterward to explain those actions.[62] For each class of capability—AI decision support, logistics autonomy, remote sensing, lethal autonomy, information operations—authoritative delegation matrices should map statutory authorities, lead agencies, and escalation routes so decision-makers and overseers know where responsibility lies.[63] For example, in the case of lethal autonomy, a named civilian signatory, such as the Under Secretary of War, must authorize actions, illustrating where accountability clearly resides. Operational systems must produce tamper-evident logs of inputs, decisions, model versions, and human actions, and agencies should set fixed interdisciplinary review windows (for example, thirty, ninety, and 180 days) so compliance and effectiveness are checked routinely, with authorized oversight bodies (inspectors general and relevant congressional committees) having pre-defined access rights.[64, 65] Acquisition language should be harmonized across agencies to require auditability, software provenance disclosure, and obligations to preserve forensic data for investigations while respecting legitimate proprietary claims.[66, 67] Contracts for critical capabilities should include clear clauses granting investigators and authorized auditors timely access to vendor artifacts—subject to appropriate protections for sensitive intellectual property—so independent reconstructions are possible after incidents.[68] Agencies should field rapid legal review teams and maintain logged legal attestations for urgent decisions so that short-notice legal advice is documented and preserved along with operational evidence.[69] High-consequence actions must pass named human authorization gates with designated civilian signatories, and degraded-communications protocols should define fallback authorities and procedures that preserve civilian oversight while enabling necessary action in the field.[70, 71] Civil-military concurrence templates should specify when State Department or White House concurrence is needed for operations with major foreign-policy implications and provide rapid timelines or clear escalation pathways if full concurrence is not feasible in the allotted time.[72] Make legal, privacy, and civil-liberties attestation a required element of system certification before fielding in interagency contexts to ensure systems meet baseline normative and statutory constraints.[73] Finally, preserve secure whistleblower channels and statutory protections so personnel can report unsafe or noncompliant fielding without

Accountability starts with clarity about who may act, when they may act, and what evidence will exist afterward to explain those actions.

retaliation, enabling internal correction and supporting external oversight when necessary.[74]

Private Sector, Allies, and Standards

Public-private partnerships are essential since commercial vendors supply cloud infrastructure, pretrained models, sensors, and perception software that government operations increasingly depend on. Agencies should adopt supplier risk-management frameworks that require vendors to demonstrate secure software-development lifecycle practices, supply-chain transparency, vulnerability disclosure processes, and commitments to preserve forensic artifacts for audits.[75] Contracts must balance proprietary interests with oversight needs by specifying clear access rights for audits, obligations to retain logs and software snapshots, and requirements for timely incident reporting.[76, 77] Maintaining a centralized government model registry that records production models, versions, provenance metadata, training data lineage, performance metrics, and authorized use cases, enabling auditors and operators to trace decisions to model artifacts and avoid unnecessary duplication. Model registries and associated metadata tools are a recognized Machine Learning Operations (MLOps) best practice for reproducibility, provenance, and governance of deployed AI systems.[78, 79] Furthermore, incorporating shared sandbox environments where suppliers can test updates rapidly while still meeting audit obligations can help balance oversight with innovation incentives and show empathy for industry partners.

Governance must be measurable and adaptive so that policies and technical controls improve as technologies and threats evolve.

Working with allies and partners to harmonize technical and normative standards—such as human-in-the-loop thresholds, attribution confidence taxonomies, and incident investigation procedures—can reduce friction in coalition operations and encourage responsible vendor behavior in multinational supply chains.[80, 81] Where appropriate, pursue mutual recognition of trusted partners' certification regimes or establish reciprocal acceptance agreements to simplify coalition fielding while preserving assurance levels and interoperable safety expectations.[82, 83] Finally, participate actively in international standards bodies (ISO/IEC), multilateral AI policy fora (OECD, G7), and coalition working groups to align technical specifications and legal norms that underpin crossborder operations and procurement.[84, 85]

Metrics, Continuous Improvement, and Transparency

Governance must be measurable and adaptive so that policies and technical controls improve as technologies and threats evolve.[86] It is also imperative to develop clear key performance indicators (KPIs) for compliance and auditability, providing oversight bodies with objective measures of governance implementation. Among the KPIs, prioritize the percentage of systems with tamper-resistant logs and the time required to produce a full audit trail during an investigation as critical indicators of accountability improvement. Additionally, track the share of deployments with pre-deployment legal attestation.[87] Including operational capability KPIs, such as time to achieve shared situational awareness, accuracy of resource assignment in crisis, and reduction in conflicting messages across agencies, is imperative to capture whether governance improvements are actually improving joint outcomes.[88]

It is also important to track ethical outcomes using measurable indicators such as the incidence of documented bias in model outputs, the number of adverse incidents attributable to autonomy

or information operations, and trends in public complaints or litigation to surface systemic harms that require remediation.[89] Importance should also be placed on rapid after-action reporting (classification permitting) that produces both internal lessons learned and public summaries where possible, and that requires those reports to feed directly into policy and acquisition updates on a short cadence.[90] Fast policy-iteration cycles should be implemented to ensure problems discovered in exercises, pilots, or incidents lead to concrete changes in standards, training, or procurement within defined timeframes.[91] Providing secure, well-publicized channels and statutory protections for whistleblowers so that personnel can report unsafe or noncompliant practices without fear of retaliation, enabling internal correction and, where appropriate, supporting external oversight.[92] To enhance organizational processes, incorporate technical safeguards such as standardized log formats, tamper-evident storage, and security controls for reporting systems in line with NIST guidance, ensuring that auditability and reporting are dependable and resilient.[93]

Tradeoffs, Tensions, and Risk Management

Improving governance inevitably introduces frictions that can slow some forms of improvisation and rapid field innovation, creating a tension between agility and assurance that must be managed explicitly.[94] Conversely, permissive or underspecified rules increase the likelihood of legal violations, erosion of public trust, and longterm harm to norms that enable coalition activity and interagency cooperation.

To manage this tradeoff, specify clear thresholds that determine when different governance regimes apply. For example, using risk tiers tied to potential harm, mission sensitivity, or likelihood of cross-border effects—so actors know when stricter certification, named civilian signatories, or broader diplomatic concurrence are required.[95] Labeling these graduated risk tiers as green for low risk, amber for moderate risk, and red for high risk can aid quick reference during operations and embed the concept in daily language.

Pre-approved playbooks and delegation matrices can preserve speed within bounded scenarios by authorizing specified actions in advance while retaining rollback and reporting requirements that protect accountability.[96, 97] A graduated tolerance approach prioritizes quick, lower-assurance deployment of humanitarian or assessment tools with low consequences, while insisting on comprehensive lifecycle certification and legal attestation for lethal or diplomatically sensitive systems. This strategy allows the interagency to accept reasonable risks when the benefits exceed the potential harms.[98, 99]

Improving governance inevitably introduces frictions that can slow some forms of improvisation and rapid field innovation...

Use red-teaming, pilots, and controlled experiments to reveal failure modes before widespread fielding, and mandate report-fast cycles so negative outcomes trigger rapid rollback, remediation, and policy updates.[100, 101] Finally, maintain clear incentives for honest reporting—secure whistleblower channels, protected afteraction reviews, and institutionalized learning mechanisms—so agencies identify and correct problems early without penalizing prudent candor.[102,103]

Conclusion

Interagency operations can be both fast and accountable if the wholeofgovernment approach is grounded in shared technical standards, transparent auditability, and embedded ethical processes that support professional judgment and civilian control.[104, 105] Practical steps include a

mandatory interagency datagovernance framework and model documentation, operational validation and continuous supervision of models, lifecycle certification and forensic logging for autonomous systems, preauthorized information playbooks and a standing joint information cell, an ethics horizon board and embedded ethics resources, clear delegation matrices, standardized acquisition clauses for vendor audit access, a centralized model registry, and a phased implementation plan built around pilots and exercises with measurable KPIs.[106, 107, 108] These measures require leadership, funding, and sustained commitment to iterative learning across agencies, with allies, and with privatesector partners, but they make it possible to preserve civilian control, uphold legal norms, and maintain professional judgment while adapting to rapidly changing operational environments.[109, 110] Democratic values—grounded in transparency, accountability, rule of law, and respect for individual rights—provide the essential framework for ensuring that the use of advanced technologies remains both legitimate and publicly accountable. By institutionalizing collective stewardship of AI, autonomy, and information tools, the interagency can retain speed without sacrificing accountability, ensuring that decisions made in contested and timepressured settings remain traceable, reviewable, and aligned with democratic values.[111]

Imagine a scenario where, when faced with a global cyber threat affecting multiple nations, this framework springs into action. Agencies from different countries share critical intelligence through the centralized model registry, swiftly aligning on the nature of the threat. Ethical boards provide rapid advisory inputs, ensuring that responses not only mitigate the immediate danger but also uphold international legal standards. As pre-authorized playbooks are deployed, collaborative teams follow clear governance steps and liaise with private vendors under standardized acquisition clauses. Multinational cooperation, bolstered by robust accountability and streamlined communication, turns the tide, ushering in a new era in which interagency operations navigate crises with both efficiency and ethical integrity. This vignette highlights the potential of a well-structured framework to not only avert a crisis but also to reinforce global alliances and public trust. ***IAJ***

Notes

1 Institute for the Study of War, "The Russian Invasion of Ukraine: Assessments and Reporting," *Institute for the Study of War,* 2022, https://www.understandingwar.org

2 U.S. Army Futures Command, Transforming in Contact (concept brief), *U.S. Army Futures Command,* 2022, https://armyfuturescommand.com (Note: Public concept materials and doctrine summaries on Transforming in Contact were released by Army Futures Command and related Army offices in 2022–2024)

3 Arjen Boin, Paul 't Hart, Eric Stern, and Bengt Sundelius, *The Politics of Crisis Management: Public Leadership under Pressure*, Cambridge University Press, 2005.

4 Louise K. Comfort, "Crisis Management in Hindsight: Cognition, Communication, Coordination, and Control," *Public Administration Review* 67, s1 (2007): 189-197, https://doi.org/10.1111/j.1540-6210.2007.00827.x.

5 National Security Commission on Artificial Intelligence, *Final Report*, 2021, https://www.dwt.com/-/media/files/blogs/artificial-intelligence-law-advisor/2021/03/nscai-final-report--2021.pdf

6 Anna Jobin, Marcello Ienca, and Effy Vayena, "The Global Landscape of AI Ethics Guidelines," *Nature Machine Intelligence* 1, no. 9 (2019): 389-399, https://doi.org/10.1038/s42256-019-0088-2.

7 Stanley A. McChrystal, *Team of Teams: New Rules of Engagement for a Complex World* (New York, NY: Portfolio/Penguin, 2015).

8 D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison, "Hidden Technical Debt in Machine Learning Systems," (presented at the Proceedings of the 2015 Workshop on Machine Learning Systems), https://proceedings.neurips.cc/paper_files/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf

9 National Security Commission on Artificial Intelligence. *Final Report*.

10 Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, B., Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru , "Model Cards for Model Reporting" (presented at the Proceedings of the Conference on Fairness, Accountability, and Transparency (FAccT), 220-229, January 29, 2019, ACM), https://doi.org/10.1145/3287560.3287596.

11 Sculley et al., "Hidden Technical Debt in Machine Learning Systems."

12 Mitchell et al., "Model Cards for Model Reporting."

13 Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy, "Explaining and Harnessing Adversarial Examples," arXiv, 2015, https://arxiv.org/abs/1412.6572.

14 Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *California Law Review* no 104 (2016): 671-732, doi:http://dx.doi.org/10.2139/ssrn.2477899

15 Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (presented at the Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 77-91, 2018).

16 Nicholas Diakopoulos, "Accountability in Algorithmic Decision Making," *Communications of the ACM* 59, no. 2 (2016): 56-62, https://doi.org/10.1145/2844110.

17 Goodfellow, Shlens, and Szegedy, "Explaining and Harnessing Adversarial Examples."

18 National Institute of Standards and Technology, *AI Risk Management Framework* (Version 1.0), U.S. Department of Commerce, 2023, https://www.nist.gov/ai-risk-management.

19 Mitchell et al., "Model Cards for Model Reporting."

20 Joaquin Quiñonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D. Lawrence, *Dataset Shift in Machine Learning* (Cambridge, MA: MIT Press, 2009).

21 National Institute of Standards and Technology, *AI Risk Management Framework* (Version 1.0).

22 U.S. Department of Defense, *DoD Directive 3000.09: Autonomy in Weapon Systems*, 2012, https://www.esd.whs.mil/Directives/issuances/dod-dir-300009/.

23 Diakopoulos, "Accountability in Algorithmic Decision Making."

24 National Institute of Standards and Technology, *AI Risk Management Framework* (Version 1.0).

25 Mitchell et al., "Model Cards for Model Reporting;" Sculley et al., "Hidden Technical Debt in Machine Learning Systems."

26 Ibid.

27 Bruno Siciliano and Oussama Khatib, eds. *Springer Handbook of Robotics*, 2nd ed. (New York City, NY: Springer, 2016).

28 Federal Aviation Administration, *Integration of Civil Unmanned Aircraft Systems (UAS) in the National Airspace System Roadmap*. U.S. Department of Transportation, 2013, https://www.faa.gov/uas/resources/roadmap/.

29 U.S. Department of Defense. *DoD Directive 3000.09: Autonomy in Weapon Systems*.

30 Ibid.

31 National Institute of Standards and Technology, *AI Risk Management Framework* (Version 1.0).

32 Ryan Calo, "Robots in American Law," (University of Washington School of Law Research Paper, 2016), https://ssrn.com/abstract=2737598

33 Federal Aviation Administration, *Unmanned Aircraft Systems (UAS) Integration in the National Airspace System: Progress and Challenges*. U.S. Department of Transportation, 2018, https://www.faa.gov/uas/resources/

34 Roland Siegwart, Illah Riza Nourbakhsh, & David Scaramuzza, *Introduction to Autonomous Mobile Robots*, 2nd ed., (Cambridge, MA: MIT Press, 2011).

35 U.S. Department of Defense. *DoD Directive 3000.09: Autonomy in Weapon Systems*.

36 Jobin et al., "The Global Landscape of AI Ethics Guidelines."

37 National Institute of Standards and Technology, *National Supply Chain Risk Management Practices for Federal Information Systems* (NIST Interagency Report 7622). U.S.

38 U.S. Government Accountability Office, *Software Acquisition: Additional Actions Needed to Help DOD Implement Future Modernization Efforts* (GAO-23-105611) (April 5, 2023), https://www.gao.gov/products/gao-23-105611

39 International Civil Aviation Organization, *Manual on Aircraft Accident and Incident Investigation* (Doc 9756), 2011.

40 B. Siciliano and O. Khatib, *Springer Handbook of Robotics.*

41 Finale DoshiVelez & Been Kim, Towards a Rigorous Science of Interpretable Machine Learning, 2017, arXiv, https://arxiv.org/abs/1702.08608

42 Robert Chesney and Danielle K. Citron, "Deepfakes and the New Disinformation War," *Foreign Affairs* 98, no. 1 (2019): 147-155, https://www.jstor.org/stable/26798018

43 US Congress House of Representatives, *Smith-Mundt Modernization Act of 2012*, HR 5736, 112th Cong., 2nd sess., introduced in House May 10, 2012, https://www.congress.gov/bill/112th-congress/house-bill/5736.

44 *War and National Defense*, U.S. Code 50 (2026), Legal Information Institute, Cornell Law School, https://www.law.cornell.edu/uscode/text/50.

45 Thomas Rid and Ben Buchanan, "Attributing Cyber Attacks," *International Security* 41, no. 4 (2015), 59-99, https://doi.org/10.1080/01402390.2014.977382

46 Patrick Meier, *Digital Humanitarians: How Big Data is Changing the Face of Humanitarian Response*, (New York, NY: Routledge, 2015).

47 European Commission, *Code of Practice on Disinformation*, 2018, https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation.

48 Federal Emergency Management Agency, *National Incident Management System (NIMS)*, U.S. Department of Homeland Security, 2017, https://www.fema.gov/sites/default/files/2020-07/fema_nims_whole.pdf.

49 E. Higgins, *Bellingcat's Guide to Open Source Investigation*. Bellingcat, 2018. https://www.bellingcat.com/resources/how-tos/2018/01/08/bellingcats-guide-open-source-investigation/

50 Rid and Buchanan, "Attributing Cyber Attacks."

51 Meier, *Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response*.

52 Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, … and Effy Vayena, "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds and Machines* 28, no. 4 (2018): 689-707, https://doi.org/10.1007/s11023-018-9482-5.

53 AI Now Institute, *AI Now 2018 Report*, AI Now Institute, 2018, https://ainowinstitute.org/AI_Now_2018_Report.pdf

54 European Commission High Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

55 Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnet Gebru… "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing" (presented at the *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (FAccT): 33-44, 2020, https://doi.org/10.1145/3351095.3372833

56 AI Now Institute, *AI Now 2018 Report*

57 McChrystal, *Team of Teams.*

58 Floridi et al., "AI4People—An Ethical Framework for a Good AI Society."

59 Raji et al., "Closing the AI Accountability Gap."

60 European Commission High Level Expert Group on Artificial Intelligence.

61 AI Now Institute. (2018). AI Now 2018 report.

62 *Inspector General Act of 1978*, Public Law 95–452, 92 Stat. 1101 (codified as amended at 5 U.S.C. App.), Council of the Inspectors General on Integrity and Efficiency, https://www.ignet.gov/content/inspector-general-act-1978.

63 Federal Emergency Management Agency, *National Response Framework*, 3rd ed. U.S. Department of Homeland Security, 2019. https://www.fema.gov/sites/default/files/2020-07/fema_national_response_framework_2019.pdf.

64 National Institute of Standards and Technology. (2006). *Guide to Computer Security Log Management* (NIST Special Publication 800-92), U.S. Department of Commerce, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-92.pdf

65 Inspector General Act of 1978, Pub. L. No. 95–452, 92 Stat. 1101 (1978). https://www.congress.gov/95/plaws/publ452/PLAW-95publ452.pdf

66 U.S. General Services Administration, "FAR Part 17 — Special Contracting Methods: Subpart 17.5 — Interagency Acquisitions," *Acquisition.gov*. https://www.acquisition.gov/far/17.5.

67 National Institute of Standards and Technology, *National Supply Chain Risk Management Practices for Federal Information Systems* (NIST Interagency Report 7622).

68 Ibid.

69 Ibid.

70 U.S. Department of Defense. *DoD Directive 3000.09: Autonomy in Weapon Systems*.

71 Federal Emergency Management Agency. *National Response Framework*.

72 [72] Ibid.

73 National Institute of Standards and Technology, *AI Risk Management Framework* (Version 1.0).

74 Whistleblower Protection Enhancement Act of 2012, Pub. L. No. 112 199, 126 Stat. 1376 (2012), https://www.congress.gov/112/plaws/publ199/PLAW-112publ199.pdf.

75 National Institute of Standards and Technology, *National Supply Chain Risk Management Practices.*

76 U.S. General Services Administration, "FAR Part 17 — Special Contracting Methods."

77 National Institute of Standards and Technology, "*National Supply Chain Risk Management."*

78 Sculley et al., "Hidden Technical Debt in Machine Learning Systems."

79 M. Zaharia, Andrew Chen, A. Davidson, A. Ghodsi, S. Hong, A. Konwinski…and Corey Zumar "Accelerating the Machine Learning Lifecycle with MLflow," IEEE Data Engineering, 2018, http://sites.computer.org/debull/A18dec/p39.pdf

80 Organisation for Economic Co-operation and Development, *OECD Principles on Artificial Intelligence*, 2019, https://www.oecd.org/going-digital/ai/principles/

81 ISO/IEC JTC 1/SC 42, *Artificial Intelligence Standards*, International Organization for Standardization / International Electrotechnical Commission, https://www.iso.org/committee/6794475.html

82 North Atlantic Treaty Organization Standardization Office, "NATO Standardization: Enabling Coalition Operations," NATO Standardization Office, 2019, https://nso.nato.int

83 European Commission, *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

84 Organisation for Economic Co-operation and Development, *OECD Principles on Artificial Intelligence*.

85 ISO/IEC JTC 1/SC 42. *Artificial Intelligence Standards*.

86 National Institute of Standards and Technology. *AI Risk Management Framework* (Version 1.0).

87 Ibid.

88 Government Performance and Results Act Modernization Act of 2010, Pub. L. No. 111352, 124 Stat. 3866 (2011), https://www.congress.gov/111/plaws/publ352/PLAW-111publ352.pdf

89 Raji et al., "Closing the AI Accountability Gap."

90 Federal Emergency Management Agency, *National Response Framework*.

91 National Institute of Standards and Technology. *AI Risk Management Framework* (Version 1.0).

92 Whistleblower Protection Enhancement Act of 2012, Pub. L. No. 112 199, 126 Stat. 1376.

93 National Institute of Standards and Technology. *Security and Privacy Controls for Information Systems and Organizations* (NIST Special Publication 800-53, Rev. 5). U.S. Department of Commerce, 2020, https://doi.org/10.6028/NIST.SP.800-53r5.

94 McChrystal, *Team of Teams.*

95 Organisation for Economic Co-operation and Development, *OECD Principles on Artificial Intelligence*.

96 Federal Emergency Management Agency, *After Action Reporting and Improvement Planning Guidance*, U.S. Department of Homeland Security, 2019, https://www.fema.gov.

97 European Commission High Level Expert Group on Artificial Intelligence.

98 U.S. Department of Defense, *DoD Directive 3000.09: Autonomy in Weapon Systems*.

99 National Institute of Standards and Technology, *AI risk management framework (Version 1.0).* U.S.

100 Sculley et al., "Hidden Technical Debt in Machine Learning Systems."

101 National Institute of Standards and Technology, *AI Risk Management Framework* (Version 1.0).

102 Whistleblower Protection Enhancement Act of 2012.

103 Federal Emergency Management Agency, *National Response Framework*.

104 National Institute of Standards and Technology, *AI Risk Management Framework* (Version 1.0).

105 Mitchell et al., "Model Cards for Model Reporting."

106 Ibid.

107 U.S. Department of Defense, *DoD Directive 3000.09: Autonomy in Weapon Systems*.

108 AI Now Institute, *AI Now 2018 Report*.

109 National Security Commission on Artificial Intelligence, *Final Report*.

110 Organisation for Economic Co-operation and Development, *OECD Principles on Artificial Intelligence*.

111 National Institute of Standards and Technology, *AI Risk Management Framework* (Version 1.0).