JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4, Issue 2, June 2026

ALGORITHMIC ATTRIBUTION: THE PROMISE AND PERIL OF AI-DRIVEN CHEMICAL FORENSICS FOR THE CHEMICAL WEAPONS CONVENTION REGIME

Abdul Rehman*

Research Officer at the Centre for International Strategic Studies AJK, Specializing in the Chemical Weapons Convention, Non-Proliferation Regimes, And Weapons of Mass Destruction

DOI: https://doi.org/10.5281/zenodo.22841452

Keywords

Article History
Received: 28 April 2026
Accepted: 15 June 2026
Published: 30 June 2026

Corresponding Author: *
Abdul Rehman

Abstract

The integration of artificial intelligence and machine learning into chemical forensics represents a transformative yet precarious shift for the verification architecture of the Chemical Weapons Convention. While computational approaches such as deep neural network spectral interpretation, non target environmental analysis, and automated impurity profiling offer unprecedented analytical speed and sensitivity, they collide directly with a treaty regime designed around human expert judgment, physical custody chains, and finite chemical schedules. This paper interrogates the legal, political, and structural consequences of allowing algorithmic models to mediate international chemical weapons attribution. The paper argues that the technical opacity of advanced machine learning models compromises established norms of scientific reproducibility and procedural fairness, while the proliferation of generative molecular design tools threatens to render traditional schedule based control lists obsolete. Furthermore, because advanced computational infrastructure remains concentrated among a handful of technologically privileged states, the uncritical deployment of these tools risks entrenching an algorithmic verification divide that could severely erode institutional trust in the Organisation for the Prohibition of Chemical Weapons. To resolve these challenges, the paper presents a comprehensive governance architecture founded on methodological explainability, multilateral cross laboratory validation, equitable capacity transfer, and the strict preservation of human interpretive authority in high stakes compliance decisions.

1. Introduction

On the morning of April 7, 2018, residents of Douma, a besieged suburb east of Damascus, reported that helicopters had dropped barrel bombs containing an agent based on chlorine. Within days the incident had metastasized into a full blown geopolitical crisis. The United States, United Kingdom, and France launched retaliatory missile strikes before any international inspectors had set foot on the ground. The Organisation for the Prohibition of Chemical Weapons (OPCW) found itself trapped in a crossfire of competing forensic narratives that it was institutionally inadequate to adjudicate. The Fact Finding Mission eventually concluded that "reactive chlorine" had likely been used, but dissenting internal memos that surfaced months later, leaked by a whistleblower who disputed the framing of the final report, exposed something deeply uncomfortable about the state of chemical weapons attribution. Even when the analytical chemistry is sound, the interpretive layer remains stubbornly and irreducibly human, and therefore contestable (Whitaker, 2020).

This episode crystallizes a problem that has only sharpened in the years since. The Chemical Weapons Convention (CWC), negotiated throughout the 1980s and opened for signature in January 1993, was designed around a verification paradigm that assumed human analysts, physical sample chains, and a relatively stable universe of known chemical agents. Its drafters, working in the long shadow of the Cold War and the chemical horrors of the war between Iran and Iraq, could scarcely have imagined a world in which deep neural networks classify mass spectrometry data faster than a human technician can load a sample, or in which generative molecular models can propose novel nerve agent analogues in a single afternoon (Urbina et al., 2022). The treaty's normative architecture is running on 20th century epistemological hardware. The forensic tools available to states and international organizations are rapidly acquiring 21st century capabilities. The gap between the two is not merely technical. It is legal, political, and deeply consequential for the credibility of the entire nonproliferation regime.

The present paper takes this gap as its central object of inquiry. It examines the intersection of artificial intelligence and machine learning with chemical forensics as applied to CWC verification, attribution, and compliance. The core question is deceptively simple: what happens when algorithmic systems begin to mediate the evidentiary processes on which international arms control depends?

The question is not hypothetical. Machine learning classifiers are already being trained on spectral libraries of scheduled chemical warfare agents and their degradation products (Ahrens et al., 2022). Non target analysis pipelines powered by deep learning can identify previously unknown compounds in environmental matrices drawn from alleged use sites (Sloop et al., 2023).Impurity profiling algorithms can, at least in principle, trace a seized nerve agent back to its synthetic route and, by extension, to the specific laboratory that produced it. These are not speculative futures. They are operational or nearly operational realities in a handful of technologically advanced national laboratories, most of them located in NATO member states.

Yet the CWC contains no provisions governing the admissibility, validation, or contestation of algorithmically generated forensic evidence. The Verification Annex, which sets out the procedures for sampling, chain of custody, and laboratory analysis, stipulates requirements that presuppose human interpretive judgment at every stage (Krutzsch et al., 2014). The OPCW network of Designated Laboratories operates under quality assurance protocols calibrated for workflows driven by human analysts. The political bodies that ultimately receive and act upon forensic findings, namely the Executive Council and the Conference of States Parties, have no established framework for evaluating the provenance or reliability of a neural network's output. This is a lacuna of considerable magnitude. It sits at the precise juncture where technical capability and institutional readiness are most misaligned.

Three research questions structure the analysis that follows. First, how are AI and machine learning currently being applied, or proposed for application, in chemical forensics relevant to CWC verification and attribution? This is a descriptive and technical question, but it carries normative weight because the specific architectures and data pipelines involved determine what kinds of evidence AI can produce and what kinds of errors it is prone to making. Second, what legal, evidentiary, and political challenges does forensic evidence generated by algorithms pose for the credibility and legitimacy of the CWC regime? Here the paper engages with the "black box" problem, meaning the opacity of many highly capable ML models, and its implications for scientific reproducibility, due process in international investigations, and the already fragile consensus around chemical weapons attribution. Third, what governance frameworks are needed to ensure that chemical forensics enhanced by AI strengthens rather than corrodes the arms control architecture? This final question is constructive and future oriented, drawing on emerging AI governance debates in adjacent domains.

The significance of this inquiry extends well beyond the technical particulars of spectral classification or molecular fingerprinting. At stake is a foundational premise of the CWC itself: the idea that objective, scientifically verifiable evidence can serve as a neutral arbiter in disputes over chemical weapons use. That premise has always been somewhat aspirational. The Douma controversy, the Syrian government's persistent denials, and Russia's instrumentalization of procedural objections within the OPCW all attest to the political malleability of even the most rigorous forensic findings (Barber, 2018).But AI introduces a qualitatively different kind of complication. When the evidentiary chain passes through an algorithm whose decision making logic may be opaque even to its own developers, the grounds for contestation multiply. States already inclined to reject unfavorable findings will find fresh ammunition. States lacking the technical capacity to independently validate claims generated by algorithms will find themselves in a position of epistemic dependency. The OPCW, already strained by the political fallout from its Syria investigations, will face pressure to either embrace tools it cannot fully audit or reject capabilities that could materially improve its investigative reach. Neither option is appealing.

This paper's contribution is twofold. On the analytical side, it bridges a disciplinary divide that has, to date, kept the machine learning literature in analytical chemistry largely insulated from the arms control and international law scholarship that grapples with CWC verification. Chemists and data scientists developing these tools rarely ask whether their models would survive scrutiny in an OPCW investigative context. Arms control scholars, for their part, have devoted considerable attention to autonomous weapons systems but comparatively little to the quieter revolution unfolding in forensic laboratories. The present analysis attempts to bring these conversations into genuine contact. On the policy side, the paper pays particular attention to the implications for states with developing technical capacities, with a specific focus on Pakistan. As a significant CWC State Party whose chemical industry profile, regional security environment, and nascent AI ecosystem intersect in distinctive ways, Pakistan offers a revealing case study in the broader dynamics of technological asymmetry within the nonproliferation regime.

Methodologically, the paper is qualitative and interdisciplinary. It draws on technical literature in analytical chemistry and machine learning, treaty law analysis of the CWC text and OPCW institutional practice, international security scholarship on verification and compliance, and policy documents produced by the OPCW Scientific Advisory Board and Conference of States Parties. I make no claim to original laboratory work or computational experimentation. The contribution is analytical and synthetic, aimed at mapping a problem space that has not yet received sustained scholarly attention.

One caveat is worth stating plainly at the outset. The field of AI in chemical forensics is moving quickly, and any snapshot of its current state risks obsolescence within a few years. The arguments advanced here are therefore less about specific algorithms or datasets than about the structural and institutional problems that will persist regardless of which particular ML architecture happens to dominate the field at any given moment. The black box problem, the verification divide, the dual use dilemma, the evidentiary gap: these are durable challenges, and they demand durable responses.

2. Literature Review: At the Intersection of Three Fields

The scholarship relevant to this inquiry sits at the confluence of three largely self-contained bodies of literature, and the most striking feature of that confluence is how little cross pollination has actually occurred. Researchers in analytical chemistry and machine learning have produced a growing corpus of work on algorithmic classification of chemical compounds, yet they rarely pause to consider the treaty law implications of deploying their models in an international verification context. Arms control scholars, meanwhile, have written extensively about CWC compliance and the politics of chemical weapons attribution without engaging the technical specifics of how AI might alter the evidentiary landscape. The emerging literature on AI governance in international security has focused overwhelmingly on autonomous weapons systems, leaving forensic and analytical applications almost entirely unexamined. This section maps these three strands and identifies the gap that the present paper seeks to fill.

The technical literature on machine learning in analytical chemistry has expanded rapidly over the past decade. Early work focused on chemo metric applications of support vector machines and random forests for spectral classification (Luts et al., 2010), but the field has since moved toward deep learning architectures capable of interpreting complex mass spectrometry and ion mobility data with minimal human preprocessing (Valdez et al., 2018).Of particular relevance to chemical forensics is the growing body of work on non-target analysis, where ML pipelines identify unknown compounds in environmental samples without relying on preexisting reference libraries (Stefanuto et al., 2021). More provocatively, Urbina and colleagues showed that a toxicity prediction model could be inverted to generate novel toxic molecules in hours, raising dual use concerns that the chemistry community has only begun to grapple with (Urbina et al., 2022). What unites this literature is its almost exclusive focus on performance metrics such as accuracy, precision, and recall. The question of whether a model's output would satisfy the evidentiary standards of an international treaty regime is conspicuously absent.

The CWC verification literature, by contrast, is richly attentive to institutional and legal detail but largely silent on computational innovation. The authoritative commentary by Krutzsch, Myjer and Trapp provides an exhaustive account of the treaty's verification mechanisms, from routine inspections under Article VI to challenge inspections under Article IX, yet its treatment of analytical methodology assumes traditional laboratory workflows(Krutzsch et al., 2014) . More recent scholarship has focused on the political dimensions of attribution, particularly following the establishment of the OPCW Investigation and Identification Team in 2018. Alexander Kelle offer a careful analysis of how the IIT's mandate has been contested by states skeptical of its findings(Kelle, 2026), while Zanders situates the broader verification crisis within the erosion of multilateral arms control norms. None of these works, however, considers how algorithmic evidence might complicate or reshape the attribution debates they describe.

The third strand, arms control theory and the politics of technical knowledge, provides the conceptual scaffolding for this paper but has not yet been applied to AI in chemical forensics. Gallagher and Chayes and Chayes established foundational frameworks for understanding how verification functions as both a technical and a political process, where "adequate" verification is always a negotiated threshold rather than an objective standard (Chayes & Chayes, 2019)(Gallagher 1999). Peter M. Haas introduced the concept of epistemic communities to explain how networks of technical experts shape international policy coordination, a framework that maps neatly onto the OPCW Scientific Advisory Board and its Designated Laboratory network (Haas, 1992). Whereas Horowitz and Garcia have begun to theorize the implications of

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4, Issue 2, June 2026

AI for international security and arms control more broadly, but their attention remains fixed on military applications rather than forensic or analytical ones(Garcia, 2025) (C. Horowitz, 2018). The gap, then, is structural rather than incidental. Each literature has developed sophisticated tools for understanding its own domain, but the intersection where AI driven chemical forensics meets treaty verification and international legitimacy remains largely uncharted. This paper positions itself squarely in that gap, drawing on all three strands to construct an integrated analysis of what algorithmic attribution means for the CWC regime and for states navigating its evolving demands.

3. Technical Foundations: How AI and Machine Learning Are Reshaping Chemical Forensics

Any serious discussion of what artificial intelligence means for the Chemical Weapons Convention must begin with a clear understanding of what these tools actually do in a laboratory setting. The temptation in policy circles is to treat "AI" as a monolithic capability, something that either works or does not, that is either trustworthy or suspect. The reality is considerably more granular. Machine learning intervenes at specific points in the forensic workflow, with specific architectures suited to specific tasks, and each intervention carries its own error profile and interpretive demands. This section unpacks those particulars for a readership that may not be steeped in analytical chemistry but needs to grasp the technical contours well enough to evaluate the governance implications that follow.

3.1 The Traditional Science and Where AI Enters

A conventional chemical forensic investigation, whether conducted under OPCW auspices or by a national laboratory, follows a sequence that has remained broadly stable for decades. Environmental or biomedical samples are collected at the site of an alleged incident, sealed, and transported under strict chain of custody protocols to a certified laboratory. There, analysts prepare the samples through extraction, derivatization, or concentration procedures tailored to the suspected agent class. The prepared samples are then introduced into analytical instruments, most commonly gas chromatography coupled with mass spectrometry (GC-MS), liquid chromatography mass spectrometry (LC-MS), or nuclear magnetic resonance (NMR) spectroscopy. The instruments produce spectral data, which trained chemists interpret by comparing observed peaks, fragmentation patterns, and retention times against reference libraries of known compounds. The final step is reporting: the analyst synthesizes the instrumental findings into a written assessment that identifies the detected substances, estimates their concentrations, and, where possible, offers inferences about their origin or method of synthesis.

Every stage of this workflow now has a potential AI counterpart, though the degree of maturity varies enormously. Sample preparation remains largely manual. The most consequential AI interventions occur at the data interpretation stage. Traditional spectral library matching relies on dot product algorithms that compare an observed spectrum against a curated database of reference spectra, returning a similarity score. This approach works well when the target compound is known and well represented in the library. It fails when the compound is novel, degraded, or present in a complex environmental matrix that generates interfering signals. Machine learning classifiers, by contrast, can learn to recognize patterns in spectral data that do not correspond to any single reference entry, enabling what is known as non-target analysis: the identification of compounds that were not specifically sought (Dowling & Meyer, 2025).

At the attribution stage, the potential of AI is even more striking, though also more speculative. Impurity profiling, which examines the trace byproducts left behind by a particular synthetic route, has long been used in forensic chemistry to link seized drugs or explosives to specific production batches. Work done by Fraga and colleagues in 2017 demonstrated that machine learning models trained on impurity profiles of organophosphorus nerve agents could distinguish between samples produced via different synthetic pathways with high accuracy(Fraga et al., 2011). If such models were validated and deployed at scale, they could in principle enable a form of chemical "fingerprinting" that traces a weaponized agent back to the facility or even the production campaign that generated it. This would represent a qualitative leap in attribution capability, moving the field from identifying *what* was used to identifying *who* made it.

3.2 Machine Learning Architectures: A Primer for the Policy Reader

A brief taxonomy of the relevant ML approaches is necessary here, not because the reader needs to build these systems but because the governance challenges differ depending on which type of model is in play.

a) Supervised learning is the most mature and widely deployed approach in chemical forensics. In supervised learning, a model is trained on labeled data, meaning a set of spectra or chromatograms that have already been identified by human experts as belonging to specific compound classes. The model learns the statistical patterns that distinguish, say, a sarin degradation product from a background pesticide signal. Common architectures include support vector machines, random forests, and, increasingly, convolutional neural networks applied to spectral images (Luts et al., 2010). Supervised models tend to be relatively interpretable: a random forest, for instance, can reveal which spectral features contributed most to a classification decision. Their limitation is that they can only recognize what they have been trained to recognize. A supervised model trained on the three CWC Schedule 1 nerve agents will not flag a novel fourth generation agent unless it happens to share enough spectral features with the training set.

b) Unsupervised learning operates without labeled data. Techniques such as principal component analysis, t-distributed stochastic neighbor embedding (t-SNE), and autoencoders cluster samples based on their inherent spectral similarities, flagging anomalies that deviate from expected patterns. This is the backbone of non target analysis. An unsupervised model might detect an unusual cluster of peaks in an environmental sample from a suspected attack site without knowing in advance what compound produced them. The interpretive burden then shifts back to the human analyst, who must determine whether the anomaly represents a genuine threat agent, a benign industrial chemical, or an instrumental artifact. Unsupervised models are powerful precisely because they are not constrained by existing knowledge, but their outputs are inherently ambiguous and require expert contextualization.

c) Deep learning, particularly deep neural networks with many hidden layers, represents the cutting edge of spectral interpretation. These models can learn highly abstract representations of chemical data, capturing nonlinear relationships that simpler architectures miss. Deep learning classifiers applied to ion mobility spectrometry data could distinguish between multiple chemical warfare agent simulants with accuracy exceeding 98 percent. The tradeoff is opacity: deep neural networks are notoriously difficult to interpret, and explaining *why* a particular classification was reached often requires specialized techniques such as SHAP values or layer wise relevance propagation, which are themselves approximations rather than definitive explanations(Molnar, 2022). This opacity is the technical root of the "black box" problem discussed in Section 4.

d) Generative models, including variational auto encoders and transformer based molecular generators, occupy a different category altogether. Rather than classifying or detecting existing compounds, they propose entirely new molecular structures. As the now famous Urbina and colleagues experiment demonstrated, a generative model originally designed to minimize drug toxicity could be reconfigured to maximize it, producing tens of thousands of novel toxic molecules in a matter of hours(Urbina et al., 2022). From a forensics standpoint, generative models are relevant because they expand the universe of potential threat agents beyond anything captured in existing CWC Schedules or reference libraries. From a proliferation standpoint, they are alarming.

3.3 Case Illustrations

Three concrete examples help ground the preceding discussion.

The first concerns impurity profiling of organophosphorus nerve agents. Fraga and colleagues at the Pacific Northwest National Laboratory developed a workflow in which GC-MS data from sarin samples produced via different synthetic routes were fed into machine learning classifiers. The models learned to associate specific impurity signatures with specific precursor chemicals and reaction conditions. In controlled experiments, the classifiers correctly matched blind samples to their source routes with accuracy above 90 percent (Fraga et al., 2011). The implications for attribution are significant: if an OPCW investigation recovered sarin residues from a munition fragment, an impurity profiling model could potentially narrow the field of possible production sources, much as metallurgical analysis of shrapnel can sometimes trace a conventional weapon to its manufacturer. The caveat is that real world samples are far messier than laboratory preparations, and environmental degradation can obscure the impurity signatures on which the models rely.

The second illustration involves environmental monitoring at alleged use sites. Stefanuto and colleagues (2021) employed comprehensive two dimensional gas chromatography coupled with time of flight mass spectrometry (GC×GC-TOFMS) and machine learning based data processing to detect CWA degradation products in soil samples. The non-target approach allowed them to identify compounds that would have been missed by conventional targeted screening, including unexpected degradation pathways that only become visible when the full complexity of the environmental matrix is computationally resolved(Gaida et al., 2023). For OPCW field missions, which often operate under time pressure and with limited sample volumes, such capabilities could dramatically improve the sensitivity and scope of environmental investigations.

The third case is the Urbina and colleagues experiment mentioned above, which deserves emphasis because it straddles the boundary between forensics and threat generation. The researchers used a commercially available generative model, MegaSyn, and simply inverted its optimization objective from minimizing toxicity to maximizing it. Within six hours, the model had generated approximately 40,000 novel molecular structures predicted to be more toxic than VX, the most lethal nerve agent in the CWC Schedule 1 inventory. Several of these structures closely resembled known CWAs; others were entirely unprecedented. The experiment was intended as a warning, and it succeeded. It demonstrated that the same computational infrastructure that could help forensic chemists identify unknown threat agents could also help malicious actors design them(Urbina et al., 2022). This dual use dynamic is explored in greater depth in Section 5.

Taken together, these illustrations reveal a field in rapid transition. The tools are real, their performance in controlled settings is impressive, and their potential to transform chemical weapons verification is substantial. But the leap from laboratory demonstration to treaty grade forensic evidence is a long one, and it passes through terrain that is as much institutional and political as it is technical. The next section examines that terrain.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal
ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4, Issue 2, June 2026

4. The Normative and Legal Landscape: Where the CWC Meets the Algorithm

The technical capabilities described in the preceding section are, in a sense, the easy part of this problem. Building a neural network that classifies nerve agent spectra with 98 percent accuracy is a formidable engineering achievement, but it is a bounded challenge with measurable outputs. Integrating that network's conclusions into a treaty verification regime that was designed for a different epistemological era is a challenge of an entirely different order, one that implicates questions of law, institutional legitimacy, political trust, and the philosophy of scientific evidence. This section interrogates what happens when the CWC's carefully constructed evidentiary architecture encounters the algorithmic outputs it was never designed to accommodate.

4.1 Evidentiary Standards under the CWC: A Framework Built for Human Judgment

The Chemical Weapons Convention and its Verification Annex establish a detailed procedural regime for the collection, handling, analysis, and reporting of chemical evidence. Part XI of the Verification Annex, which governs investigations of alleged use, stipulates that samples must be collected by designated inspection teams, sealed in the presence of host state representatives, and transported to the OPCW laboratory or to one of the network of Designated Laboratories under unbroken chain of custody (*Part XI – Investigations in Cases of Alleged Use of Chemical Weapons*, n.d.). The analytical methods employed must conform to "recommended operating procedures" developed by the OPCW, and results must be validated through proficiency testing before a laboratory is authorized to conduct official analyses.

What is striking about this framework, when read from the vantage point of 2026, is the degree to which it presupposes a human analyst at every interpretive juncture. The Verification Annex speaks of "analysis," "identification," and "confirmation" as activities performed by qualified personnel. The quality assurance protocols developed by the OPCW Technical Secretariat assume that analytical results can be traced back to specific methodological choices made by identifiable individuals. The proficiency testing regime evaluates whether a laboratory's human staff can correctly identify blind samples using validated instrumental methods. The entire edifice rests on an implicit premise: that forensic evidence is produced by people, using instruments as tools, and that the interpretive judgment of those people is both the source of the evidence's authority and the locus of its accountability.

An AI model disrupts this premise in ways that are subtle but profound. Consider a scenario in which a Designated Laboratory employs a deep learning classifier to identify an unknown compound in an environmental sample collected from a suspected chemical attack site. The model returns a classification with a confidence score of 96 percent, identifying the compound as a degradation product of a Schedule 1 nerve agent. The human analyst reviews the output, finds it plausible, and includes it in the laboratory report. Who, in this chain, produced the evidentiary finding? The analyst who commissioned the model? The data scientists who trained it? The curators of the spectral database on which it was trained? The Verification Annex has no vocabulary for parsing this distribution of epistemic labor.

The problem is not merely semantic. It has concrete legal implications. Under the CWC, States Parties have the right to challenge inspection findings and to demand clarification of the analytical methods used. If a state disputes a conclusion derived from an AI model, what level of methodological transparency is it entitled to? Can it demand access to the training data, the model architecture, and the hyper parameter settings? Can it request that an independent laboratory reproduce the result using the same model? The treaty is silent on all of these questions, and the silence is not benign. It creates

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4, Issue 2, June 2026

a zone of ambiguity that states with adversarial interests will inevitably exploit.

4.2 The Black Box Problem and Scientific Reproducibility

The opacity of many high performing machine learning models constitutes perhaps the most fundamental tension between AI driven forensics and the norms of international scientific verification. In the natural sciences, reproducibility is the bedrock of epistemic legitimacy. A finding that cannot be independently replicated by a competent laboratory using the same methods is, by definition, suspect. The OPCW's reliance on a network of Designated Laboratories reflects this principle. When the Technical Secretariat needs to confirm a finding, it can distribute samples to multiple independent facilities and compare results.

Deep neural networks complicate this picture considerably. A convolutional neural network trained on mass spectrometry data may achieve extraordinary classification accuracy, but the internal representations it learns are distributed across thousands or millions of parameters in ways that resist straightforward interpretation. The model "knows" that a particular spectral pattern corresponds to a sarin degradation product, but it cannot explain its reasoning in the language of chemical structure and fragmentation pathways that a human analyst would use to justify the same conclusion. This is the so called black box problem, and it is not a temporary limitation that will be solved by the next generation of architectures. It is, to a significant degree, an inherent feature of the most powerful models currently available (Molnar, 2022).

The implications for CWC verification are troubling. Imagine that the OPCW Fact Finding Mission relies on an AI assisted analysis to conclude that a particular chemical agent was used in an attack. A member state disputes the finding and requests that the analysis be replicated by a different Designated Laboratory. If the second laboratory lacks access to the same trained model, or if it uses a model with a different architecture trained on a different dataset, it may arrive at a different conclusion even when analyzing the same sample. Is this a failure of reproducibility? Or is it simply a reflection of the fact that different models, like different human analysts, may interpret ambiguous data differently? The CWC framework has no mechanism for adjudicating between these possibilities.

The field of explainable AI (XAI) offers partial remedies. Techniques such as SHAP values, attention maps, and layer wise relevance propagation can generate post hoc explanations of a model's decisions, highlighting which input features contributed most to a given classification (Molnar, 2022). These tools are genuinely useful, but they are approximations rather than faithful reconstructions of the model's internal logic. A SHAP explanation can tell you that a particular mass to charge ratio was influential in the model's classification, but it cannot tell you whether the model's reasoning is chemically sound or whether it has latched onto a spurious correlation in the training data. For a forensic finding that may trigger diplomatic crises or military responses, approximations are not enough.

There is also the question of adversarial robustness. Machine learning models are notoriously vulnerable to adversarial perturbations: small, carefully crafted modifications to input data that cause the model to produce wildly incorrect outputs while remaining imperceptible to human observers (Goodfellow et al., 2015). In a chemical forensics context, this raises the unsettling possibility that a sophisticated actor could deliberately contaminate a sample in a way that causes an AI classifier to misidentify the agent, either to conceal the use of a prohibited substance or to fabricate evidence of use where none occurred. The CWC's chain of custody protocols are designed to prevent physical tampering, but they were not designed to guard against adversarial manipulation of computational pipelines.

4.3 Can the OPCW Absorb AI?

The question of whether the OPCW as an institution can integrate AI driven tools into its verification and investigative workflows is as much a question of resources, expertise, and political will as it is of technical feasibility. The organization operates on a modest budget by the standards of international organizations, and its Technical Secretariat, while staffed by highly qualified chemists and engineers, does not currently include specialists in machine learning or data science . The OPCW Chemtech Centre in Hague, which serves as the central analytical hub for the organization, is equipped with state of the art instrumentation but is recently trying to deploy AI models in its routine analytical workflows.

The Scientific Advisory Board (SAB) has shown awareness of the issue. Its reports to the Fifth Review Conference in 2023 flagged artificial intelligence and machine learning as areas of emerging relevance to the Convention, noting that these technologies could "significantly enhance the speed and accuracy of chemical identification" while also raising "questions about validation, transparency, and the potential for misuse". The Scientific Advisor Boards Temporary Working Groups has launched two important reports in 2026.First on the subject of Artificial Intelligence and the other on Chemical Forensics(Artifical Intelligence, 2026) (Chemical Forensics, 2026). These reports are ground breaking work but the affective implementation of their recommendations is still a question which OPCW has to cope with. This is a useful starting point, but the SAB's recommendations are advisory rather than binding, and the gap between acknowledging a technological trend and developing the institutional infrastructure to govern it is substantial.

The Designated Laboratory network presents a particular challenge. As of 2026, the network comprises approximately two 34 laboratories, the vast majority of which are located in Europe and North America (*Designated Laboratories*, n.d.). These laboratories vary considerably in their technical sophistication, and the introduction of AI tools would almost certainly widen the existing capability gap. A laboratory in the Netherlands or Sweden with access to high performance computing infrastructure and trained data scientists could deploy deep learning classifiers that a laboratory in a developing country simply could not replicate. This asymmetry has direct implications for the reproducibility and perceived fairness of OPCW investigations, a point explored further in Section 6.

The political bodies of the OPCW face their own adaptation challenges. The Executive Council and the Conference of States Parties are composed of diplomats, not data scientists, and their capacity to evaluate the reliability of AI generated forensic findings is limited. When the IIT presented its reports on chemical weapons use in Syria, the debates in the Executive Council focused overwhelmingly on the political implications of the findings rather than on the technical methodologies underpinning them. Introducing algorithmic evidence into this already fraught political environment would add a layer of complexity that many delegations are poorly equipped to navigate.

4.4 Precedent and Practice: Lessons from the Syria Investigations

The OPCW's investigations into alleged chemical weapons use in Syria provide the most relevant real world precedent for thinking about how AI driven forensics might function, or malfunction, in a contested political environment. Between 2014 and 2023, the Fact Finding Mission and the Investigation and Identification Team conducted numerous investigations into incidents involving chlorine, sarin, and sulfur mustard, producing a body of forensic evidence that was simultaneously rigorous in its analytical methodology and fiercely contested in its political reception (Pita & Domingo, 2026)

The Douma investigation of 2018, discussed briefly in the introduction, is particularly instructive. The FFM's conclusion that reactive chlorine had been used was based on a combination of environmental sampling, biomedical analysis, and witness testimony, all processed through conventional analytical methods. Yet the finding was challenged by Russia and Syria on grounds that included alleged inconsistencies in the sampling methodology and the handling of witness accounts. The leaked internal dissent, in which an OPCW engineer questioned whether the final report accurately reflected the raw data, further eroded confidence in the investigative process among skeptics (Farell, 2019) .

Now consider how this episode might have unfolded differently if AI driven forensics had been part of the investigative toolkit. On one hand, machine learning models could potentially have identified trace compounds in the environmental samples that human analysts missed, strengthening the evidentiary basis for the FFM's conclusions. Non target analysis algorithms might have detected degradation products that conventional targeted screening overlooked, closing gaps that critics later exploited. On the other hand, the introduction of algorithmic evidence would have given skeptics an entirely new line of attack. Russia, which has consistently opposed the IIT's mandate, could have challenged the transparency and reproducibility of any AI generated findings, demanding access to model architectures and training data that the OPCW might be unable or unwilling to provide. The black box problem would have been weaponized, quite literally, in the service of political obstruction.

This is not a hypothetical concern. The broader pattern of the Syria investigations demonstrates that forensic evidence in the CWC context is never evaluated in a political vacuum. Every analytical finding is filtered through the strategic interests of the states involved, and the credibility of the evidence is inseparable from the credibility of the institution producing it. AI tools that enhance analytical sensitivity but reduce interpretive transparency would, in this environment, be a double edged sword. They could make the OPCW's investigations more technically powerful while simultaneously making them more politically vulnerable.

The lesson for the CWC regime is that the integration of AI into chemical forensics cannot be treated as a purely technical upgrade. It is a normative intervention that reshapes the evidentiary foundations on which the entire verification architecture rests. Getting the governance right is not a secondary concern to be addressed after the technology has been deployed. It is a precondition for deployment that must be addressed now, before the gap between capability and regulation becomes unbridgeable.

5. The Dual Use Dilemma: AI as Sword and Shield

There is an uncomfortable symmetry at the heart of this entire inquiry, one that no amount of technical optimism can wish away. The same computational architectures that enable a forensic chemist to identify an unknown nerve agent in a soil sample from a Syrian village can, with relatively minor modifications, enable a malicious actor to design a novel toxic compound that no existing detection system would recognize. This is not a theoretical risk. It has already been demonstrated in a peer reviewed journal, by researchers who intended their work as a warning and inadvertently produced a proof of concept for chemical weapons proliferation in the age of artificial intelligence. The dual use character of AI in the chemical sciences is not a side effect or an afterthought. It is the central ethical and security problem that any governance framework must confront.

5.1 Generative Chemistry and the De Novo Design of Toxic Molecules

The experiment that brought this problem into sharp public focus was published in March 2022 by Fabio Urbina and his colleagues at Collaborations Pharmaceuticals, a small drug discovery firm based in North Carolina (Urbina et al., 2022).The researchers were working with a machine learning model called MegaSyn, which had been trained on molecular datasets to predict the toxicity of pharmaceutical compounds. The model's intended purpose was benign: by identifying toxic molecular features early in the drug design process, it could help medicinal chemists avoid pursuing candidates that would prove harmful in clinical trials. This is a routine application of AI in the pharmaceutical industry, and dozens of similar models are in active commercial use.

What Urbina and his team did was conceptually simple but deeply unsettling. They inverted the model's optimization function. Instead of penalizing toxicity, they rewarded it. They then asked the model to generate novel molecular structures that would maximize predicted lethality while remaining within the bounds of chemical plausibility. The results were startling. Within six hours of continuous computation, the model had generated approximately 40,000 unique molecular structures, many of which were predicted to be more toxic than VX, the most lethal nerve agent currently listed in Schedule 1 of the CWC (Urbina et al., 2022). Some of the generated molecules bore close structural resemblance to known chemical warfare agents. Others were entirely unprecedented, occupying regions of chemical space that no human chemist had previously explored. The researchers did not synthesize any of these compounds, and they emphasized in their publication that the experiment was intended to provoke discussion about the risks of accessible AI tools in the chemical domain. The discussion it provoked was intense and remains unresolved.

The significance of this experiment for the CWC regime is difficult to overstate. The Convention's control architecture is built around three Schedules of chemicals, which enumerate specific compounds and classes of precursors subject to declaration, inspection, and restriction. Schedule 1 covers the most dangerous agents, including sarin, VX, sulfur mustard, and their immediate precursors. Schedule 2 and Schedule 3 cover less lethal but still militarily relevant chemicals and their precursors. The Schedules were negotiated in the late 1980s and early 1990s based on the known chemical weapons landscape of that era. They are, by design, finite lists of known substances.

Generative AI models expose a fundamental vulnerability in this list based approach. If a machine learning system can propose tens of thousands of novel toxic molecules that fall outside the existing Schedules, then the CWC's control regime is, in principle, perpetually one step behind the frontier of computational chemistry. A state or non-state actor equipped with generative models could theoretically design a chemical weapon that is functionally equivalent to a Schedule 1 agent but structurally distinct enough to evade both the treaty's legal definitions and the reference libraries used by forensic laboratories. This is what some scholars have begun to call the "Schedule Zero" problem: the growing universe of toxic chemicals that are not scheduled but that pose equivalent threats (Revill et al., 2022).

The CWC does contain a general purpose criterion in Article II, which defines a chemical weapon not by its molecular structure but by its intended use. Under this criterion, any toxic chemical produced for hostile purposes is a chemical weapon regardless of whether it appears on a Schedule. In theory, this closes the loophole. In practice, the general purpose criterion is extraordinarily difficult to enforce. Proving intent requires intelligence and attribution capabilities that go well beyond chemical analysis, and the evidentiary burden in a contested political environment is immense. A novel compound generated by an AI model and synthesized in a small laboratory would be virtually undetectable by routine OPCW inspections unless inspectors knew precisely what to look for, which they would not.

5.2 AI Enabled Proliferation Risks Beyond Molecular Design

The generative chemistry problem, alarming as it is, represents only one facet of a broader proliferation landscape shaped by AI. Several other developments deserve attention.

The first concerns automated synthesis planning. Tools such as IBM RXN for Chemistry and Synthia (formerly Chematica) use machine learning to propose synthetic routes for target molecules, effectively automating a task that traditionally required years of graduate training in organic chemistry (Segler et al., 2018). These tools are commercially available and increasingly sophisticated. In a benign context, they accelerate drug development and materials science. In a malign context, they could lower the knowledge barrier for producing chemical warfare agents, enabling actors with limited chemistry expertise to identify viable synthesis pathways for toxic compounds. The combination of generative molecular design and automated synthesis planning creates a pipeline that moves from concept to producible compound with minimal human chemical knowledge, a prospect that should give arms control policymakers considerable pause.

The second development involves the democratization of AI capabilities through open source models and cloud computing. The machine learning models used in the Urbina experiment were not proprietary military systems. They were built on publicly available frameworks and trained on datasets accessible through commercial chemical databases. As AI tools become cheaper, faster, and more user friendly, the technical threshold for conducting the kind of experiment Urbina and his colleagues performed continues to fall. A well-funded non state actor with access to cloud computing resources and basic data science expertise could, in principle, replicate and extend the Collaborations Pharmaceuticals experiment without requiring a state level research infrastructure (Lentzos, 2022).

The third concern relates to the intersection of AI with synthetic biology and biochemistry. While this paper focuses on chemical weapons, the boundary between chemical and biological agents is increasingly porous. AI models trained on protein structure prediction, such as DeepMind's AlphaFold, have revolutionized structural biology and opened new avenues for designing novel toxins and bio regulators (Jumper et al., 2021). The convergence of AI driven chemistry and AI driven biology could produce hybrid threats that fall between the mandates of the neither CWC and nor the Biological Weapons Convention, exploiting the jurisdictional gaps between the two regimes.

5.3 Governance Responses: The Hague Ethical Guidelines and Beyond

The international community is not entirely without tools for addressing these risks, though the existing tools are blunt and largely voluntary. The most relevant framework is the set of Hague Ethical Guidelines, adopted by the OPCW Conference of States Parties in 2015 and reviewed at the Fifth Review Conference in 2023. The Guidelines articulate a set of principles for responsible conduct in the chemical sciences, emphasizing the obligation of chemists to prevent the misuse of their knowledge and calling on states, industry, and academia to promote a culture of responsibility. They are a valuable normative statement, but they were drafted before the advent of generative AI and do not specifically address computational tools or machine learning models.

The question of whether the Hague Ethical Guidelines should be updated to encompass AI is now actively debated within the OPCW's Scientific Advisory Board and among member states. Several proposals have been floated. One approach would be to extend the Guidelines to cover the development and deployment of AI models in the chemical sciences, requiring researchers to conduct dual use risk assessments before publishing work that could facilitate the design of toxic compounds. Another would be to establish a voluntary code of conduct for AI developers working with chemical data, analogous to the responsible disclosure practices used in cybersecurity. A more ambitious proposal would

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4, Issue 2, June 2026

involve creating an international registry of AI models trained on chemical warfare agent data, subjecting them to a form of export control similar to the Wassenaar Arrangement's provisions for dual use technologies (Revill et al., 2022).

Each of these proposals faces significant implementation challenges. Voluntary codes of conduct are difficult to enforce and easy to ignore. Risk assessment requirements could stifle legitimate research and drive sensitive work underground. An international registry would require a level of transparency that both states and private companies are unlikely to accept. And the pace of AI development means that any regulatory framework risks obsolescence within a few years of adoption.

The deeper problem is structural. The CWC was designed to control physical substances: specific chemicals, specific precursors, and specific production facilities. AI operates in the domain of information and computation, which are inherently more difficult to monitor, restrict, and verify. You can inspect a chemical plant. You cannot easily inspect a neural network running on a cloud server in a jurisdiction that does not cooperate with international inspectors. This asymmetry between the material focus of the treaty and the informational nature of the threat represents a fundamental challenge that no incremental update to existing frameworks can fully resolve.

What is needed, and what the following sections begin to sketch, is a more comprehensive approach to AI governance in the chemical weapons domain. Such an approach would need to address not only the technical capabilities of AI models but also the institutional, legal, and political structures that determine how those capabilities are deployed, validated, and contested. The dual use dilemma cannot be eliminated. It can only be managed, and the quality of that management will depend on the willingness of states, international organizations, and the scientific community to confront uncomfortable tradeoffs between innovation, security, and transparency.

6. The Verification Divide: Equity, Capacity, and Geopolitical Implications

The distribution of AI capability in chemical forensics is not random. It maps almost perfectly onto existing patterns of scientific and economic privilege, and this geographic concentration carries consequences that extend well beyond the laboratory. If the tools that determine whether a state is found to have used chemical weapons are developed, trained, and controlled by a small circle of technologically advanced nations, then the perceived neutrality of the entire CWC verification regime comes under strain. This section examines what I call the algorithmic verification divide and its implications for the legitimacy and durability of chemical arms control.

6.1 Who Builds the Algorithms?

A survey of the peer reviewed literature on machine learning applications in chemical threat detection reveals a striking geographic skew. The overwhelming majority of published research originates from laboratories in the United States, the United Kingdom, the Netherlands, Germany, Sweden, Finland, and the Czech Republic. These are, not coincidentally, the same states that host the most capable OPCW Designated Laboratories and that contribute the largest shares of the organization's budget and technical personnel. The OPCW network of Designated Laboratories, which currently numbers approximately 34, is concentrated in Europe and North America. Laboratories in the Global South remain underrepresented, and those that do participate often lack the computational infrastructure and specialized personnel required to develop or validate AI driven analytical methods.

This concentration matters because machine learning models are not neutral instruments. They reflect the data on which they were trained, the design choices of their developers, and the institutional priorities of the organizations that fund them. A classifier trained primarily on spectral data from European and North American

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4, Issue 2, June 2026

reference collections may perform poorly when confronted with compounds or environmental matrices that are more common in South Asian, African, or Latin American contexts. The training data problem is well documented in other AI domains, from facial recognition to medical diagnostics, where models trained on demographically narrow datasets produce systematically biased outputs when deployed on diverse population. There is no reason to believe that chemical forensics is immune to analogous distortions.

6.2 The Algorithmic Verification Divide and Arms Control Legitimacy

The concept of an algorithmic verification divide draws on a longer tradition of arms control scholarship that has examined how technical asymmetries shape the politics of compliance. Nancy W. Gallagher demonstrated that verification is never a purely technical exercise but rather a negotiated process in which the capacity to produce and interpret evidence is itself a source of political power(Gallagher, 1999).Abraham Chayes and Antonio H. Chayes argued that compliance with international agreements depends less on coercive enforcement than on the perceived legitimacy of the compliance process. When states believe that the evidentiary standards applied to them are fair, transparent, and accessible, they are more likely to accept unfavorable findings. When they perceive those standards as rigged or opaque, they are more likely to defect (Chayes & Chayes, 2019).

The introduction of AI tools into CWC verification risks widening the gap between states that can produce algorithmic evidence and states that can only receive it. Consider a scenario in which an OPCW investigation relies on a deep learning model developed at a European Designated Laboratory to identify a novel compound in samples collected from a Middle Eastern or South Asian site. The state under investigation lacks the technical capacity to independently evaluate the model's architecture, training data, or error profile. It must either accept the finding on faith or reject it on principle. Neither outcome strengthens the regime. Acceptance without understanding breeds resentment. Rejection without grounds breeds paralysis.

States already skeptical of the OPCW's impartiality would find in the algorithmic verification divide a powerful new rhetorical weapon. Russia's persistent objections to the IIT's Syria findings have centered on procedural and methodological concerns, many of them exaggerated or fabricated but some grounded in genuine ambiguities in the investigative process. The opacity of AI models would multiply such ambiguities exponentially. A state seeking to obstruct an investigation need not disprove the forensic evidence; it need only cast doubt on the algorithm that produced it, demanding levels of transparency that the OPCW may be unable to provide without compromising proprietary or classified information.

6.3 Capacity Building and the Article XI Gap

Article XI of the CWC commits States Parties to promoting the exchange of chemicals, equipment, and scientific information for peaceful purposes and to fostering economic and technological development in the chemical field. The OPCW's International Cooperation and Assistance Division runs a range of capacity building programs, including analytical chemistry training courses, laboratory twinning arrangements, and equipment donations to National Authorities in developing countries. These programs have been valuable in building baseline analytical capabilities across the Global South, but they have not yet incorporated AI or machine learning into their curricula.

This omission is increasingly difficult to justify. If AI driven forensics becomes the standard for CWC verification within the next decade, as many technical experts anticipate, then states that lack the capacity to develop, deploy, or critically evaluate these tools will find themselves in a position of permanent epistemic dependency.

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4, Issue 2, June 2026

They will be subject to findings they cannot independently verify, produced by models they cannot audit, using training data they cannot access. This is a recipe for the erosion of the multilateral consensus that underpins the Convention.

The remedy is not simply to export AI tools from the Global North to the Global South, though technology transfer is part of the solution. What is needed is a more fundamental investment in building indigenous AI capacity within National Authorities and Designated Laboratories in developing countries. This means training data scientists alongside chemists, investing in computational infrastructure, and ensuring that the reference libraries and spectral databases used to train AI models include data from diverse geographic and industrial contexts. The OPCW's capacity building programs should be expanded to include dedicated modules on machine learning for chemical analysis, and the Scientific Advisory Board should be tasked with developing validation protocols that are accessible to laboratories with limited computational resources.

7. Towards a Governance Framework: Responsible Integration of AI into Chemical Weapons Verification

The preceding sections have painted a picture of considerable complexity. AI driven chemical forensics offers genuine and substantial improvements in detection sensitivity, analytical speed, and attribution precision. It also introduces profound challenges to evidentiary transparency, institutional legitimacy, and the equitable distribution of verification capacity. The dual use potential of generative models adds a layer of risk that the CWC's existing control architecture is poorly equipped to manage. The question is no longer whether these technologies will reshape the chemical weapons landscape but whether the international community can develop governance structures fast enough to ensure that the reshaping is constructive rather than corrosive. This section proposes a framework organized around five core principles and three institutional mechanisms, drawing on lessons from adjacent domains of AI governance and arms control practice.

7.1 Five Principles for Algorithmic Verification

Any governance framework for AI in CWC verification must begin with a set of normative commitments that can command broad consensus among States Parties. I propose five principles, each grounded in the specific tensions identified in earlier sections.

I. **Transparency of method:** AI models used in OPCW mandated investigations must be accompanied by documentation that explains, in terms accessible to a technically literate but non specialist audience, the model's architecture, training data provenance, performance characteristics, and known failure modes. This does not require the disclosure of proprietary source code, which would be commercially and politically impractical, but it does require a level of methodological openness that exceeds current norms in the machine learning community. The OPCW Technical Secretariat should develop standardized reporting templates for AI assisted analyses, modeled on the existing recommended operating procedures for conventional instrumental methods.

II. **Reproducibility across laboratories:** A forensic finding generated by an AI model must be independently verifiable by any Designated Laboratory in the OPCW network, not merely by the laboratory that developed the model. This implies that trained models, or at minimum the training data and architectural specifications necessary to reproduce them, must be shared among Designated Laboratories under appropriate security protocols. The current practice of individual laboratories developing proprietary analytical pipelines is incompatible with the treaty's commitment to multilateral verification.

III. **Inclusivity of access:** AI tools deployed under OPCW auspices must be accessible to all States Parties, regardless of their domestic

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal
ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4, Issue 2, June 2026

technological capacity. This principle follows directly from the analysis of the verification divide in Section 6. If AI driven forensics becomes the de facto standard for CWC investigations while only a handful of states possess the infrastructure to evaluate or challenge the results, the regime's legitimacy will erode among the majority of its membership.

IV. Complementarity rather than substitution: AI should augment human expert judgment, not replace it. The interpretive layer of chemical forensics, the stage at which analytical findings are contextualized within the broader investigative narrative, must remain under human control. This principle is consistent with the broader consensus in AI ethics that high stakes decisions should not be fully automated. In the CWC context, it means that AI generated classifications should be treated as inputs to a human deliberative process rather than as self-sufficient evidentiary conclusions.

V. Accountability for errors: Clear protocols must establish who bears responsibility when an AI assisted analysis produces an incorrect or misleading result. The distributed nature of AI development, involving data curators, model architects, training engineers, and end users, makes accountability harder to assign than in traditional forensic workflows. The OPCW should develop explicit liability frameworks that specify the obligations of each actor in the AI analytical chain.

7.2 Three Institutional Mechanisms

Principles without institutional machinery are aspirational at best. Three concrete mechanisms would help translate the foregoing commitments into operational practice.

First, the OPCW Scientific Advisory Board should be expanded to include members with demonstrated expertise in machine learning, data science, and AI ethics. The SAB's current composition is heavily weighted toward traditional analytical chemistry and chemical engineering, which is appropriate for its existing mandate but insufficient for the challenges ahead. A dedicated SAB working group on computational methods in chemical verification should be established, tasked with evaluating AI tools before they are deployed in official investigations and with developing validation standards that account for the specific characteristics of algorithmic evidence.

Second, the OPCW should create a shared AI validation platform accessible to all Designated Laboratories. This platform would host standardized benchmark datasets, pre trained models, and evaluation protocols, enabling laboratories to test and compare AI tools under controlled conditions. The platform would serve a function analogous to the existing proficiency testing regime but adapted for computational rather than purely instrumental methods. Funding could be drawn from the OPCW's voluntary contributions framework, with targeted support from technologically advanced States Parties committed to Article XI capacity building.

Third, AI governance should be formally integrated into the CWC Review Conference cycle. The five yearly review process provides the most appropriate venue for States Parties to assess the impact of emerging technologies on the Convention's operation and to negotiate updates to verification protocols. The Fifth Review Conference in 2023 and the subsequent conferences took initial steps in this direction by acknowledging AI as a relevant development, but future review conferences should move beyond acknowledgment to adopt specific decisions governing the use of algorithmic evidence in OPCW investigations.

7.3 Broader Implications

The governance challenges identified here are not unique to the CWC. The Biological Weapons Convention faces analogous, and in some respects more acute, problems as AI driven tools transform bioscience research and pathogen surveillance (Revill et al., 2022). The Non Proliferation Treaty regime is beginning to grapple with the implications of machine learning for nuclear safeguards verification. A cross regime dialogue on AI in arms control verification, perhaps convened under the auspices of the United Nations Office for Disarmament Affairs, could help identify common principles and avoid duplicative institutional efforts. The window for proactive governance is narrowing. The technologies are advancing faster than the treaties designed to constrain them, and the cost of inaction will be measured not in abstract institutional credibility but in the concrete failure to prevent or attribute the next chemical weapons attack.

8. Conclusion

The Chemical Weapons Convention was built on a deceptively simple wager: that scientific verification could substitute for political suspicion. If states could trust the evidence, they would not need to trust each other. For three decades, that wager has largely held, sustained by the painstaking labor of analytical chemists working within a framework of shared protocols and mutual accountability. The advent of artificial intelligence in chemical forensics now tests that wager in ways its architects could not have foreseen.

This paper has argued that AI driven forensic tools represent both the most promising and the most perilous development in CWC verification since the treaty's entry into force. The promise is real. Machine learning classifiers can identify threat agents in complex environmental matrices with a speed and sensitivity that surpass human capability. Impurity profiling algorithms offer the tantalizing prospect of tracing a chemical weapon to its source. Non target analysis pipelines can detect compounds that conventional screening would miss entirely. These are not speculative possibilities. They are operational realities in a handful of advanced laboratories, and their diffusion is accelerating.

The perils are equally real and considerably less well understood. The opacity of deep learning models collides with the CWC's foundational commitment to reproducible, contestable evidence. The geographic concentration of AI capability threatens to create a two tier verification regime in which some states produce algorithmic findings and others merely receive them. The dual use potential of generative models has already been demonstrated in peer reviewed literature, exposing the fragility of a control architecture built around finite lists of scheduled chemicals. The institutional infrastructure of the OPCW remains largely unprepared for the integration of computational tools into its core workflows.

The governance framework proposed in Section 8 is not a panacea. Transparency, reproducibility, inclusivity, complementarity, and accountability are necessary conditions for responsible AI integration, but they are not sufficient. Realizing them will require sustained political will, significant financial investment, and a willingness among technologically advanced states to share capabilities they might prefer to hoard. The window for proactive governance is narrow. Once AI driven forensics becomes embedded in OPCW practice without adequate safeguards, retrofitting governance will be far more difficult than building it from the outset.

Future research should move beyond the analytical mapping attempted here. Empirical testing of AI models against OPCW proficiency test data, comparative studies of AI governance across arms control regimes, and field assessments of National Authority readiness in developing countries would all advance the conversation from principle to practice. The stakes could hardly be higher. In a world already struggling with the erosion of multilateral norms, the integrity of chemical weapons verification is not a technical luxury. It is a geopolitical necessity.

Bibliography

Ahrens, A., Allers, M., Bock, H., Hitzemen, M., & Ficks, A. (2022). Detection of Chemical Warfare Agents with a Miniaturized High-Performance Drift Tube Ion Mobility Spectrometer Using High-Energetic Photons for Ionization. *Analytical*

Chemistry, *94*(44). https://pubs.acs.org/ancham/article/94/44/15440/848386/Detection-of-Chemical-Warfare-Agents-with-a

Artifical Intelligence. (2026). OPCW.

Barber, R. (2018). Uniting for Peace Not Aggression: Responding to Chemical Weapons in Syria Without Breaking the Law. *Journal of Conflict and Security Law*, *24*(1), 71–110. https://doi.org/10.1093/jcsl/kry029

C. Horowitz, M. (2018). Artificial Intelligence, International Competition, and the Balance of Power. *Texas National Security Review*, *1*(2). https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/

Chayes, A., & Chayes, A. H. (2019). On compliance. *International Organization*, *47*(2), 175–205. https://doi.org/10.1017/S0020818300027910

Chemical Forensics. (2026). OPCW.

Designated Laboratories. (n.d.). OPCW. Retrieved September 9, 2026, from https://www.opcw.org/designated-laboratories

Dowling, G. M., & Meyer, M. R. (2025). Current trends in targeted and non-targeted metabolomics in analytical toxicology. *Frontiers in Molecular Biosciences*, *12*. https://doi.org/10.3389/fmolb.2025.1614399

Farell, J. A. (2019, December 27). *OPCW Douma Docs-Release Part 4*. Wiki Leaks. https://wikileaks.org/opcw-douma/

Fraga, C. G., Acosta, G. A. P., Crenshaw, M. D., Wallace, K., Mong, G. M., & Colburn, H. A. (2011). Impurity profiling to match a nerve agent to its precursor source for chemical forensics applications. *Analytical Chemistry*, *83*(24), 9564–9572. https://doi.org/10.1021/ac202340u

Gaida, M., Stefanuto, P.-H., & Focant, J.-F. (2023). Theoretical modeling and machine learning-based data processing workflows in comprehensive two-dimensional gas chromatography-A review. *Journal of Chromatography.* A, *1711*. https://doi.org/10.1016/j.chroma.2023.464467

Gallagher, N. W. (1999). *The Politics of Verification*. Johns Hopkins University Press.

Garcia, D. (2025). The global diplomacy of governing military artificial intelligence. *Ethics and Information Technology*, *27*. https://link.springer.com/article/10.1007/s10676-025-09863-0

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). *Explaining and harnesing Adverserial Examples*. ICLR.

Haas, P. M. (1992). Introduction: Epistemic Communities and International Policy Coordination. *International Organization*, *46*(1), 1–35.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kelle, A. (2026). *Norm Contestation and Compliance in the Chemical and Biological Weapons Prohibition Regimes*. CBWNet.

Krutzsch, W., Myjer, E. P. J., Trapp, R., & Herbach, J. (Eds.). (2014). *The Chemical weapons convention: A commentary*. Oxford University Press.

Lentzos, F. (2022, March 22). *Artificial intelligence could be repurposed to create new biochemical weapons*. King's College London. https://www.kcl.ac.uk/news/artificial-intelligence-could-be-repurposed-to-create-new-biochemical-weapons

Luts, J., Ojeda, F., Van de Plas, R., De Moor, B., Van Huffel, S., & Suykens, J. A. K. (2010). A tutorial on support vector machine-based methods for classification problems in chemometrics. *Analytica Chimica Acta*, *665*(2), 129–145. https://doi.org/10.1016/j.aca.2010.03.030

Molnar, C. (2022). *Interpretable Machine Learning*.

Part XI – Investigations in cases of Alleged Use of Chemical Weapons. (n.d.). OPCW. Retrieved September 9, 2026, from https://www.opcw.org/chemical-weapons-convention/annexes/verification-annex/part-xi-investigations-cases-alleged-use

Pita, R., & Domingo, J. (2026). The Use of Chemical Weapons in the Syrian Conflict. *Toxics*. https://doi.org/10.3390/toxics2030391

Revill, J., Garzón Maceda, M., & Zhang, V. (2022). *Stakeholder perspectives on the Biological Weapons Convention*. UNIDIR. https://doi.org/10.37559/WMD/22/BWC/03

Segler, M. H. S., Pruess, M., & P. Waller, M. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*. https://doi.org/10.1038/nature25978

Sloop, J., Chao, A., & Gunderson, J. (2023). Demonstrating the Use of Non-targeted Analysis for Identification of Unknown Chemicals in Rapid Response Scenarios. *Envirnmental Science and Technology*, *57*(8). https://www.researchgate.net/publication/368574216_Demonstrating_the_Use_of_Non-targeted_Analysis_for_Identification_of_Unknown_Chemicals_in_Rapid_Response_Scenarios

Urbina, F., Lentzos, F., Invernizzi, C., & Ekins, S. (2022). Dual use of artificial-intelligence-powered drug discovery. *Nature Machine Intelligence*, 4(3), 189–191. https://doi.org/10.1038/s42256-022-00465-9

Valdez, C. A., Leif, R. N., Hok, S., & Hart, B. R. (2018). Analysis of chemical warfare agents by gas chromatography-mass spectrometry: Methods for their direct detection and derivatization approaches for the analysis of their degradation products. *Reviews in Analytical Chemistry*, *37*(1), 20170007. https://doi.org/10.1515/revac-2017-0007

Whitaker, B. (2020, January 2). The Douma chemical weapons investigation and the role of Ian Henderson. *Medium*. https://brian-whit.medium.com/the-douma-chemical-weapons-investigation-and-the-role-of-ian-henderson-ab44399a9070