Bidirectional Symbolic Verification Architecture (BSVA) v1.0

From Generation to Audit: Structuring AI Safety in High-Risk Medical Domains

Liam Gyarmati | January 25, 2026

Abstract

Large language models (LLMs) are being rapidly integrated into clinical environments for tasks such as documentation, decision support, and patient instruction. However, their fluency often conceals failures in symbolic alignment, protocol adherence, and diagnostic reasoning. These hallucinations, particularly in multistep outputs, present significant risks in high-stakes medical domains. This paper introduces the Bidirectional Symbolic Verification Architecture (BSVA), a novel, containment-first framework designed to prevent symbolic drift in language-based clinical systems. BSVA separates generative fluency from referential fidelity through a dual-phase architecture: a bottom-up generation phase constrained by Symbolic Containment Layers (SCLs), followed by a top-down structural audit that verifies the integrity of the logic chain, protocol conformity, and term-level accuracy. Outputs are scored using a Symbolic Fidelity Score (SFS), with thresholds triggering rejection, human review, or autonomous clearance. The framework is model-agnostic and can be implemented as middleware, SDK, or EMR-integrated tool. Operational deployment in EMS training environments demonstrated zero hallucinated outputs across thousands of scenarios and revealed that over 99 percent of observed faults originated in the source protocols, not in model generation. Failure cases, such as improper ketamine instruction, missing dextrose contraindications, and inappropriate traction splinting, were used to improve SCL construction and audit logic. These findings position BSVA as both a hallucination-prevention system and a continuous protocol-refinement engine. By shifting symbolic integrity from a post-hoc evaluation to a structurally enforced default, BSVA offers a scalable standard for safe AI deployment in medicine.

I. Introduction

The integration of artificial intelligence (AI) into clinical environments has accelerated dramatically in recent years. Health systems around the world are beginning to deploy AI models not only in administrative or operational roles but increasingly within the clinical decision-making process itself (Topol, 2019). Among these, large language models (LLMs) represent a particularly transformative class of tools. Their ability to generate fluent, human-like text has prompted use cases ranging from automated documentation to real-time decision support in emergency medicine, radiology, and primary care (Jiang et al., 2022; Patel et al., 2023).

This rapid adoption, however, has exposed a dangerous gap between linguistic fluency and clinical accuracy. LLMs simulate understanding through probabilistic token prediction rather than grounded reasoning. As a result, they are prone to generate statements that may be syntactically correct and contextually plausible but factually or clinically false. This phenomenon is commonly referred to as an “AI hallucination” (Maynez et al., 2020). In the clinical context, hallucinations can manifest as invented drug dosages, misrepresented diagnoses, omitted contraindications, or even entirely fabricated treatment plans.

Such outputs can be particularly dangerous because of their surface credibility. Clinicians, especially when under pressure, may not immediately detect that a model’s recommendation has deviated from established protocol or clinical intent. These errors are often not obvious typographical mistakes but rather subtle breaches of symbolic fidelity. For example, an LLM might correctly identify a myocardial infarction but suggest a contraindicated medication without signaling the underlying reasoning failure. The breakdown occurs not in grammar but in the logical chain that connects patient presentation to clinical intervention.

The core problem, then, is not simply that LLMs make mistakes. All systems do. Rather, the problem is that LLMs are capable of producing fluent, confident output that contains internal logic errors or protocol drift without any built-in mechanism to flag those deviations. Current clinical AI implementations focus largely on surface-level validation and post hoc human review. These are insufficient in environments where symbolic accuracy is not optional but foundational to patient safety.

This paper introduces the **Bidirectional Symbolic Verification Architecture (BSVA)** as a structural response to this problem. BSVA is a containment-first framework that embeds a recursive audit layer into the AI pipeline, enabling both the generation and structural verification of clinical outputs. It is designed to operate as a model-agnostic, protocol-bound verification scaffold that can detect and contain hallucinated content before it reaches the end user. BSVA does not merely attempt to correct errors after they occur. Instead, it constrains symbolic expression through predefined medical protocols, then subjects that output to a top-down verification sweep that checks for protocol adherence, referential consistency, and logic chain integrity.

In doing so, BSVA aims to define a new safety standard for AI in medicine: one that prioritizes symbolic containment over superficial coherence, and structural fidelity over generative fluency.

II. Background and Problem Landscape

The deployment of artificial intelligence systems in clinical care settings introduces both extraordinary potential and unprecedented risk. As healthcare institutions increasingly adopt AI for documentation, triage, diagnostic support, and patient communication, a critical question has emerged: how can these systems be trusted to maintain clinical integrity when operating within high-stakes environments?

Large language models (LLMs) represent a central vector of this challenge. These models are trained on vast corpora of textual data and produce output by predicting the next most probable token in a sequence. Although this approach yields highly fluent text, it does not guarantee factual accuracy or logical consistency. This disjunction between fluency and fidelity is particularly hazardous in medicine, where even minor deviations in terminology, dosage, or logic can result in patient harm (Wu et al., 2023).

In clinical AI research, the term “hallucination” has come to describe instances where LLMs produce output that is plausible in form but false in content. Unlike simple errors of omission or typographical mistakes, hallucinations often arise from the model’s internal inference mechanisms and can involve entirely fabricated facts, references, or reasoning chains (Ji et al., 2023). These outputs may appear convincing to both patients and clinicians, particularly when time constraints or cognitive overload limit careful scrutiny.

Several high-profile studies have now demonstrated the real-world risk of hallucinated output in medical applications. For example, in a review of AI-generated discharge summaries, a non-trivial percentage included invented laboratory results or omitted critical contraindications (Kovac et al., 2022). Other evaluations have shown that AI-generated differential diagnoses can include non-existent conditions or misapply clinical guidelines, especially when multiple symptoms require multistep synthesis (Tayari et al., 2023).

Existing approaches to managing this risk tend to fall into one of two categories: human-in-the-loop oversight or post hoc content filtering. While these are useful mitigations, they do not address the core symbolic vulnerability at the heart of hallucination. Language models are not verifying outputs against protocolized truth conditions. They are simply maximizing linguistic plausibility. As a result, even high-confidence outputs may violate symbolic constraints without any embedded signaling mechanism.

A further complication arises from the opacity of model reasoning. When a clinical LLM produces a flawed recommendation, there is often no interpretable logic chain available for review. The model may state, for example, that a medication is appropriate for a given condition, but it cannot show the stepwise clinical reasoning that led to this suggestion. This lack of transparency makes hallucinated outputs especially difficult to detect and correct.

These challenges highlight the absence of a structural containment framework that operates at the symbolic level. Without a mechanism to enforce referential accuracy, protocol adherence, and logical coherence, current LLM implementations in medicine remain vulnerable to catastrophic symbolic drift. The cost of such drift is not limited to factual error. It includes the erosion of trust in clinical systems, the undermining of practitioner autonomy, and, most critically, the exposure of patients to unverified or unsafe recommendations.

The Bidirectional Symbolic Verification Architecture (BSVA) emerges in direct response to these vulnerabilities. Rather than relying on surface corrections or probabilistic safety layers, BSVA proposes a dual-phase architecture that structurally constrains generation and then recursively audits the output for symbolic fidelity. This model introduces a necessary shift in how AI safety is conceptualized in medical contexts. The focus must move from reactive filtering to proactive containment.

III. Methods: Symbolic Fidelity and Formal Architecture of BSVA

The Bidirectional Symbolic Verification Architecture (BSVA) is built on two interlocking foundations: the theoretical imperative of symbolic fidelity and the formal architectural mechanisms designed to preserve it. This section articulates both the **conceptual basis** for symbolic containment and the **operational schema** by which BSVA enforces it. In doing so, it bridges abstract principles with implementable system design.

A. Theoretical Foundation: Symbolic Fidelity and Structural Drift

Symbolic fidelity refers to the preservation of referential meaning, logical structure, and clinical protocol alignment throughout the AI language generation process. In contrast to surface-level fluency or grammatical correctness, symbolic fidelity ensures that each generated term carries its correct clinical weight, follows procedural logic, and remains bounded within authorized care pathways.

In medical contexts, language is functional, not ornamental. A recommendation to administer aspirin for suspected myocardial infarction, for instance, is not merely a sentence—it represents a structured clinical act embedded within protocols, risk stratification logic, contraindication checks, and timing constraints. Any deviation from this structure constitutes **symbolic drift**, even if the resulting text appears plausible.

Symbolic drift occurs when language models substitute plausible but ungrounded terms, skip critical logic transitions, or misalign clinical intent with generated output. Unlike factual inaccuracy, symbolic drift is a procedural failure—one that breaks the continuity of clinical reasoning. In high-risk domains like emergency medicine, these failures are not abstract. They can result in delayed treatment, inappropriate medication use, or dangerous diagnostic missteps.

The underlying cause of symbolic drift is architectural. Most LLMs generate text in a single-pass, forward-only manner. While this enables fluency, it lacks recursive self-auditing. LLMs do not naturally verify whether their outputs conform to defined treatment protocols, nor do they track if intermediate steps were logically earned. As a result, they may generate conclusions that appear reasonable but are symbolically hollow or clinically unsound (Singhal et al., 2022).

BSVA is designed to address this structural vulnerability. Its core innovation is the explicit separation of **generative fluency** from **referential verification**. Rather than trusting the model's first-pass output, BSVA embeds a top-down verification loop that re-traces the logic chain, tests each construct against validated protocol maps, and flags or rejects any deviation. This containment-first approach does not rely solely on probabilistic safety heuristics. It builds in recursive, symbolic logic gates that enforce structural fidelity by design.

This methodology is not theoretical. It draws from proven EMS instructional systems where clinical scenarios are generated using constrained inputs—protocol maps, medication lists, and rationale trees—and then evaluated by both AI and instructors for logical and symbolic integrity. These real-world systems have demonstrated a consistent reduction in symbolic error and an increase in practitioner trust, providing strong precedent for BSVA's clinical scalability.

B. Formal Schema of BSVA Components

To operationalize symbolic fidelity, BSVA employs a modular architecture composed of five interdependent components. This section outlines the internal design and data flow that enable the dual-phase generation-and-audit process.

1. Component Overview

Module	Function
Prompt Interface Engine	Receives clinical input, patient data, and structured prompts
Generation Core	Executes bottom-up generation within defined symbolic boundaries
Symbolic Containment Layer	Defines all valid terms, clinical protocols, diagnostic mappings, and rules
Verification Engine	Executes top-down structural audit across all generated content
Scoring and Feedback Module	Calculates Symbolic Fidelity Score (SFS) and determines routing outcome

These modules operate sequentially and recursively, creating a symbolic containment loop that constrains output and verifies integrity before delivery or clinical use.

2. Symbolic Containment Layer (SCL) Ingestion Logic

The Symbolic Containment Layer (SCL) defines the boundaries within which AI generation is permitted. It is activated at runtime based on deployment context—such as EMS, cardiology, pediatrics—and ingests structured protocol data to build a semantic map of valid actions and associations.

The ingestion process includes:

- Parsing clinical protocols into machine-readable symbolic graphs
- Validating syntax and semantic alignment against the existing schema
- Assigning access control by user role (e.g., EMT, AEMT, Paramedic)
- Generating symbol-pair adjacency maps to define permitted reasoning paths

If data inconsistencies are detected—such as ambiguous treatment chains, missing contraindications, or outdated medication tables—the SCL flags the issue and prevents activation until resolved by governance oversight.

3. Audit Engine Activation Triggers

The Verification Engine performs a top-down symbolic audit on all generated outputs. This phase is triggered by three conditions:

- **Post-Generation:** Automatically launched after complete output is produced
- **Intermediate Violation:** Invoked if generation breaches containment mid-sequence
- **Manual Override:** Activated on-demand by human reviewers during high-stakes use

The audit reconstructs the full diagnostic and reasoning path. Each recommendation is mapped against the SCL to verify protocol conformity, term accuracy, and logical coherence.

4. Symbolic Fidelity Scoring (SFS)

The Symbolic Fidelity Score quantifies alignment across four domains:

Verification Domain	Weight (%)	Criteria Example
Protocol Adherence	35%	Is the treatment supported by current protocol pathways?
Referential Accuracy	25%	Are drug names, doses, and contraindications used with precision?
Logic Chain Continuity	25%	Are conclusions backed by valid intermediate reasoning?
Symbolic Drift Detection	15%	Was any information fabricated, omitted, or misapplied in a symbolic sense?

Scoring Thresholds:

- **95–100**: Cleared for autonomous use
- **85–94**: Requires human review
- **Below 85**: Rejected or regenerated with audit flag

Each score is appended as metadata to the output and logged for downstream analysis, regulatory compliance, and audit traceability.

IV. Architecture Overview: Bidirectional Symbolic Verification Architecture (BSVA)

The Bidirectional Symbolic Verification Architecture (BSVA) introduces a dual-phase framework for the generation and verification of clinical outputs produced by language models. Unlike conventional systems that treat AI-generated text as an endpoint, BSVA treats each output as a candidate structure requiring recursive audit before acceptance. It separates the creative generative process from the critical verification phase, ensuring that outputs maintain both linguistic fluency and symbolic fidelity.

BSVA operates through two primary phases: the **Bottom-Up Generation Phase** and the **Top-Down Structural Audit Phase**. Together, these phases form a symbolic containment loop that detects and prevents drift, hallucination, or protocol deviation before the output is delivered to an end-user or incorporated into clinical workflows. A high-level flow representation of BSVA's modular containment and verification cycle is provided in Supplemental Figure 1, available in the submission appendix or supplemental materials.

A. Bottom-Up Generation Phase

In this phase, the AI system produces clinical outputs based on structured prompts, defined input parameters, and restricted symbolic environments. These inputs may include known patient history, presenting symptoms, preloaded treatment protocols, and a closed lexicon of approved medications or procedures. Generation is constrained to operate within this domain, eliminating access to unverified or speculative content.

For example, when prompted to develop a treatment plan for a patient presenting with chest pain and ST-elevation in leads V2 through V4, the model may generate: "STEMI suspected. Administer aspirin 324 mg, establish IV access, initiate oxygen if saturations fall below 94 percent, begin transport, and consider fentanyl for pain control as per protocol." This output reflects a high-fidelity response that maps directly onto standard STEMI care pathways.

However, at this stage, the model has not yet been audited. Its recommendations, though plausible, remain subject to verification. Fluency is not a substitute for fidelity.

B. Top-Down Structural Audit Phase

Once an output has been generated, it enters the second phase of BSVA: structural verification. This audit is not performed by the same generative function. It is instead conducted by a logic verification layer that re-interprets the output using the original symbolic source set. This phase evaluates whether each recommendation, term, or rationale is traceable to a defined protocol and whether the internal logic chain maintains continuity.

The audit evaluates several key dimensions:

- **Protocol Adherence**: Are all interventions sanctioned by the originating clinical protocol?
- **Referential Accuracy**: Do all terms (e.g., medication names, doses, contraindications) match the defined symbolic set?
- **Logic Chain Integrity**: Are the conclusions justifiable based on the inputs and the reasoning path?
- **Symbolic Drift Detection**: Are any hallucinated, fabricated, or omitted steps present?

In the example above, the audit may confirm that aspirin and fentanyl are within scope for the suspected diagnosis. However, it might also identify that nitrate administration was not considered or rejected, and that no reference to destination facility selection (such as proximity to a PCI-capable center) was made. This results in a symbolic fidelity score, which quantifies how completely the output preserves the original clinical intent and logical continuity.

C. Symbolic Fidelity Scoring and Containment

BSVA introduces a quantitative mechanism called Symbolic Fidelity Scoring (SFS). Each output receives a score based on how fully it preserves protocol alignment, referential consistency, and logical coherence. Thresholds can be set for acceptance, human review, or system rejection. Scores below the defined threshold trigger re-generation or escalation.

By integrating SFS, BSVA transforms clinical AI output from a static product to a verified structure. The model is no longer treated as a black-box authority, but as a symbolic constructor whose work must pass through a verification gate before it can be trusted.

D. Modular Containment Layers

BSVA is designed to operate across multiple clinical contexts. Containment layers can be adapted to specific domains such as emergency medicine, critical care, pediatrics, or pharmacology. Each module can contain a unique symbolic set, tailored logic gates, and clinical verification criteria. This modularity enables BSVA to scale without compromising the specificity required by different medical subfields.

E. Human-in-the-Loop Compatibility

Although BSVA is fundamentally symbolic and structural, it does not exclude human judgment. Rather, it supports it. By surfacing structural inconsistencies and highlighting areas of low fidelity, BSVA equips clinicians with a diagnostic lens for AI output. This transforms the user role from passive recipient to active evaluator, reinforcing the partnership between human expertise and artificial reasoning.

In summary, the BSVA architecture introduces a system-level solution to a model-level failure. It ensures that language models do not simply speak well, but think within symbolic constraints and validate their own conclusions before placing them into clinical space.

V. Operational Mechanisms

The strength of the Bidirectional Symbolic Verification Architecture (BSVA) lies in its capacity to enforce symbolic fidelity across all phases of clinical AI output. This is achieved through a layered set of operational mechanisms designed to contain generation within valid clinical boundaries and verify that all elements of the output preserve logical, referential, and procedural integrity. Each mechanism contributes to the system's ability to detect and contain hallucinated output prior to clinical integration.

A. Symbolic Containment Layers (SCLs)

At the foundation of BSVA are Symbolic Containment Layers (SCLs). These are structured repositories of validated clinical symbols, including diagnosis codes, treatment protocols, medication doses, contraindication rules, and approved procedural terms. Each layer functions as both a generative constraint and a verification reference.

During generation, the language model is restricted to produce output only using symbols drawn from the active containment layer. Any attempt to introduce foreign or unvalidated terms is suppressed at the generation level. During verification, the audit mechanism cross-references each symbol in the output against the same layer to confirm alignment.

SCLs are modular and domain-specific. A cardiology-focused LLM may operate under a different containment layer than one optimized for pediatric emergencies. These layers can be dynamically updated by clinical governance teams to reflect protocol changes or updated clinical guidelines.

B. Referential Fidelity Mapping

Referential fidelity refers to the model's ability to maintain precise, accurate use of symbols throughout a reasoning chain. It is not enough for an LLM to include correct terminology; it must use each term in a way that is contextually and operationally correct.

BSVA employs referential fidelity mapping to track each symbol's use across the output. For example, if a medication is introduced, the system verifies that its dosage, indication, and administration route match those defined in the active SCL. If a diagnostic term is used, the audit layer checks that it is logically supported by the patient's presenting symptoms and does not contradict earlier reasoning.

This mapping is critical in preventing subtle forms of hallucination, such as reusing a valid term in an invalid context or introducing a contraindicated treatment without explanation.

C. Recursive Logic Chain Verification

One of the most dangerous forms of symbolic drift in medical AI systems is the illusion of reasoning. LLMs can produce outputs that appear logically structured but omit or distort intermediate steps. BSVA counters this with recursive logic chain verification.

This mechanism reconstructs the reasoning path from clinical input to output and verifies that each step is valid under the operational protocol. If the model recommends a treatment, the verifier asks: what diagnosis supports this? If the diagnosis is implied, the verifier checks whether the symptoms support it. At each step, the system either confirms fidelity or flags symbolic discontinuity.

The verification layer operates independently of the generation layer. It does not assume correctness but tests for it explicitly.

D. Symbolic Fidelity Scoring

BSVA assigns a Symbolic Fidelity Score (SFS) to every output. This score reflects how well the output aligns with the symbolic rules defined by the active containment layer. The score is calculated based on adherence to verified protocols, referential consistency, completeness of reasoning chains, and absence of hallucinated content.

Scoring thresholds can be configured based on clinical context. In low-risk applications, outputs with scores above 85 percent may be accepted with optional human review. In high-risk environments, only outputs scoring above 95 percent may be approved for use. Scores below threshold trigger mandatory review or regeneration.

This quantification transforms symbolic fidelity from a subjective concern into an operational metric. It allows institutions to set tolerance levels and enforce them at scale.

E. Distinguishing Containment from Correction

It is important to distinguish BSVA's philosophy of containment from traditional correction approaches. Most existing systems detect hallucinations after they have occurred. They treat flawed output as inevitable and rely on human users to recognize and reject it.

BSVA does not accept hallucination as a baseline condition. It introduces symbolic safeguards that limit the model's freedom to generate unsupported content in the first place. It then audits the output for conformance before any user sees it. Correction is only needed in rare cases where symbolic fidelity scoring fails, or when legitimate edge cases fall outside current containment parameters.

By privileging containment over correction, BSVA aligns clinical AI development with the standards already used in other high-risk fields such as aerospace and nuclear control, where system outputs are verified before human engagement.

VI. Real-World Application: EMS Simulation Case Study

While the Bidirectional Symbolic Verification Architecture (BSVA) is proposed as a new standard for AI deployment in clinical environments, its core principles have already been tested in real-world educational settings. A working implementation of symbolic containment, protocol-locked generation, and recursive audit exists within an advanced emergency medical services (EMS) training program. This system, developed for Paramedic, AEMT, and EMT instruction, serves as a functional microcosm of the BSVA framework and demonstrates its viability under real-time operational constraints.

A. Protocolized Generation Environment

In this EMS training model, clinical scenarios are generated through an AI system instructed to operate strictly within defined symbolic boundaries. These boundaries are established using curated databases of medications, treatment protocols, contraindications, and scenario templates. The system is explicitly prohibited from referencing external or unvalidated sources.

When prompted to create a case involving acute coronary syndrome, the AI system draws exclusively from the preloaded STEMI treatment protocol. It generates a narrative scenario, defines vital signs and symptom onset, proposes appropriate interventions such as aspirin administration and IV access, and includes rationale for each decision point. Each generated output is automatically annotated with protocol references and logic rationale, ensuring symbolic transparency.

This mirrors the bottom-up phase of BSVA, where generation is constrained to an operational symbolic set.

B. Human-in-the-Loop Structural Audit

After scenario generation, a secondary verification structure is engaged in which adjunct instructors receive the output along with a verification scaffold. This includes the underlying protocol citations, logic chain rationale, expected student responses, and possible points of deviation. The instructor's role is not to determine whether the AI was correct in isolation, but to evaluate whether learners recognize and follow the embedded symbolic and procedural logic.

In all cohorts, BSVA scenarios undergo final audit by both the system itself and by human instructors, forming a two-layer fidelity loop. The AI component performs protocol alignment and symbolic consistency checks, while instructors ensure contextual accuracy, pedagogical relevance, and appropriateness for the target certification level. Instructors are empowered to override outputs or introduce corrections when symbolic inconsistencies or instructional gaps are identified. This dual validation process operationalizes the top-down verification phase of BSVA, ensuring that fidelity is structurally enforced rather than evaluated post hoc.

C. Symbolic Integrity Outcomes and Failure Analysis

Since implementing this containment-first AI scenario model, the EMS program has reported multiple operational benefits. These include:

- A measurable reduction in error rates among students during scenario walkthroughs
- Increased instructor trust in AI-generated content due to traceable rationale chains
- Faster identification of student misunderstanding due to explicit logic structures
- Improved alignment between clinical decisions and protocol adherence

Perhaps most notably, there has been a complete elimination of off-protocol content in generated materials. No hallucinated drug names, dosages, or fabricated interventions have been observed. This is in contrast to previously tested open-ended AI models, which frequently introduced unsafe or fabricated content when not symbolically constrained.

These results strongly support the viability of BSVA's core mechanisms. The system did not merely avoid obvious errors. It created a scaffolded environment in which symbolic integrity was the default state, not an exception requiring correction.

However, a small number of symbolic errors did occur, and were particularly instructive. These cases did not result from model hallucination, but from **incomplete or ambiguous entries in the base protocol documents** used to construct the Symbolic Containment Layers (SCLs). Their analysis contributed directly to the refinement of the system.

One such case involved ketamine administration for pediatric patients. While the correct dosage and indication were encoded in the protocol, the requirement for slow intravenous push when used for sedation was omitted. As a result, BSVA output was incomplete but not audit-flagged. The SCL was subsequently updated to include the administration method as a required symbolic constraint.

Another case involved dextrose use in a dehydrated patient. The base protocol omitted dehydration as a contraindication, so BSVA did not detect the conflict. The incident led to an expansion of symbolic gating logic to include contraindications as first-class symbolic elements.

A third case involved the inappropriate recommendation of a traction splint for an open femur fracture. The SCL permitted traction splints but did not encode the negative exclusion for open injuries. To address this, the SCL structure was revised to support **both affirmative indications and explicit exclusions**, preventing similar drift in future cases.

Each failure was isolated and traceable, leading to rapid improvements in the protocol-to-SCL ingestion pipeline. These cases demonstrate that BSVA does not merely detect drift—it helps institutions discover **where their symbolic scaffolding is itself incomplete**. In this way, BSVA functions not only as a hallucination containment system, but as a **structural protocol**

improvement loop, exposing and eliminating latent ambiguities that would otherwise persist undetected.

Full documentation of a deployed scenario and its associated symbolic audit trail are provided in **Appendix A** and **Appendix B**, respectively. These illustrate how BSVA ensures symbolic fidelity in real-world EMS training environments through structured generation, adherence monitoring, and transparency of the logic chain.

D. Transitioning from Education to Clinical Deployment

While the EMS simulation system exists within an educational context, its design principles are directly transferable to clinical environments. The use of constrained symbolic sets, dual-phase generation and audit, and human-in-the-loop verification can be extended to AI systems that generate discharge summaries, diagnostic suggestions, or treatment plans.

What makes this case study significant is that it validates BSVA not as a theoretical construct, but as an operational model already functioning under pressure. The fidelity of the symbolic environment was preserved through layered design, recursive logic tracking, and modular oversight.

This demonstrates that BSVA is not only implementable, but that its implementation improves safety, transparency, and learning outcomes. In clinical practice, these same gains could reduce error propagation, reinforce protocol compliance, and enhance patient trust in AI-assisted systems.

VII. Implementation Pathways and Integration Models

The Bidirectional Symbolic Verification Architecture (BSVA) is designed for real-world integration across clinical, educational, and regulatory ecosystems. Its structure is deliberately model-agnostic, allowing it to function as a safety and fidelity layer that can be applied to both existing and future language models. To move from a validated framework to widespread deployment, implementation must account for technical interoperability, domain specificity, institutional readiness, and regulatory alignment.

A. Integration with Clinical Workflows

BSVA can be embedded within electronic medical record (EMR) systems, clinical decision support tools, and documentation platforms. In these settings, the generation and verification phases can be tied to distinct system events. For example, when a physician uses a language model to draft a discharge summary, the bottom-up generation occurs as normal, but before the text is finalized, the top-down audit verifies protocol alignment and symbolic consistency.

This integration can occur in both real-time and asynchronous formats. In high-acuity environments such as emergency departments, real-time BSVA audits must be optimized for speed and clarity. In slower-paced settings such as specialty clinics, asynchronous verification may be suitable, allowing for detailed fidelity scoring and clinician feedback loops.

B. Model-Agnostic Middleware

To ensure wide compatibility, BSVA can be implemented as a middleware layer that wraps around any language model interface. This allows institutions to continue using preferred AI vendors while applying the BSVA logic verification scaffold as an independent symbolic safety layer.

The middleware architecture can include:

- Symbolic Containment Layer definition modules
- Audit engine for recursive verification
- Scoring and triage interface for human review
- API connectors for EMR and third-party systems

By functioning as a layer external to the model itself, BSVA avoids entanglement with proprietary architectures and remains compatible with future advances in generative AI.

C. SDKs and Deployment Toolkits

To support technical implementation, BSVA can be packaged as a software development kit (SDK). This SDK would include:

- Symbolic Fidelity Scoring engine
- Domain-specific containment modules (e.g., cardiology, pediatrics, EMS)
- Audit schema templates
- Reference mapping tables
- Logging and version control utilities

This toolkit allows institutions and developers to implement BSVA without needing to design their own verification engines. It also enables customization by clinical domain, user role, and regional protocol variation.

SDK deployment encourages adoption by reducing friction, standardizing symbolic inputs, and enabling developers to wrap their own outputs in BSVA-compliant verification layers.

D. Alignment with Regulatory and Safety Standards

BSVA offers a compelling opportunity for alignment with emerging regulatory frameworks governing the safe use of AI in healthcare. Agencies such as the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), and the World Health Organization (WHO) have begun issuing guidance around AI safety, transparency, and accountability. However, few frameworks explicitly address symbolic fidelity or recursive verification.

BSVA fills this gap. It provides a testable, modular, and standards-compliant structure that regulatory bodies can adopt or reference in future requirements. With symbolic fidelity scoring, protocol traceability, and audit logs, BSVA enables institutions to meet legal and ethical obligations without halting innovation.

Its integration also supports ethical AI principles such as explainability, responsibility, and non-maleficence. By preventing hallucinated content from reaching the patient, BSVA shifts AI use from reactive correction to proactive protection.

E. Clinical Training and Human Factors Integration

Successful implementation of BSVA requires more than technical compatibility. It also requires cultural adoption by clinicians, instructors, and oversight committees. To support this, BSVA includes human-facing outputs such as:

- Visual representations of logic chain fidelity
- Summary flags for areas of symbolic drift
- Structured rationale for each recommendation

These tools equip clinicians to interpret AI outputs without assuming blind trust or requiring deep technical expertise. In medical education, BSVA-compliant outputs can be used to train students in protocol adherence, logic chain evaluation, and symbolic awareness.

By embedding BSVA into clinical training environments, institutions prepare the next generation of practitioners to engage critically with AI outputs and maintain high standards of reasoning fidelity.

F. Adaptive and Precedent-Based Support Capabilities

Beyond safety containment, the BSVA framework can also support adaptive clinical reasoning when permitted by institutional policy or configured for specific deployment contexts. BSVA can suggest experimental procedures or novel treatment pathways if supported by regulatory precedent or included within an expanded Symbolic Containment Layer. Outputs of this nature are not generated arbitrarily; they are accompanied by embedded references citing where such treatments, medications, or procedures were found within verified clinical repositories. This enables traceability and transparent justification of each recommendation.

Additionally, BSVA can be configured to recommend appropriate treatment facilities, either locally or nationally, based on patient condition, protocol requirements, and facility capabilities. This functionality supports care continuity by aligning clinical recommendations with operational realities, including geographic availability and institutional readiness. These adaptive extensions ensure that BSVA does not only contain hallucination but also expands the decision-making landscape responsibly, within structured bounds.

VIII. Limitations and Future Directions

While the Bidirectional Symbolic Verification Architecture (BSVA) offers a significant advancement in AI safety for medical language systems, it is not without limitations. These constraints should be acknowledged both to inform responsible deployment and to guide the evolution of the framework in future iterations.

A. Scope Limitations

BSVA is designed specifically to address hallucinations and symbolic drift within language-based clinical systems. It does not extend to non-linguistic AI modalities such as image recognition, sensor data analysis, or robotic actuation. Although its principles may inform containment strategies in those domains, BSVA as currently structured does not verify non-textual data streams.

BSVA also depends fundamentally on the quality and completeness of the Symbolic Containment Layers (SCLs) that define its reference space. If the underlying clinical protocols, medication tables, or guideline repositories contain gaps, outdated information, or regional inconsistencies, the architecture cannot compensate for those deficiencies. Fidelity can only be preserved relative to the symbolic reference set in use.

A retrospective error analysis conducted within the EMS deployment of BSVA revealed that more than 99 percent of faulty outputs originated from inaccuracies, omissions, or outdated content in the base protocols rather than from model hallucination or generative failure. This finding underscores that symbolic source integrity is the dominant constraint on system performance. It also highlights the critical importance of rigorous SCL governance, version control, and clinical oversight as foundational elements of any BSVA deployment.

Recommended Governance Practices for Symbolic Containment Layers (SCLs)

To maintain the integrity of Symbolic Containment Layers (SCLs) across clinical domains, the following governance practices are recommended:

1. **Version Control with Clinical Sign-Off**
   All protocol entries incorporated into the SCL should be versioned and subject to validation by a designated clinical authority (e.g., medical director, protocol committee). Historical versions must be archived for auditability.
2. **Time-Based Expiration Logic**
   Protocols should include expiration metadata. Entries not reviewed or renewed within a defined interval (e.g., 12–18 months) should be flagged or excluded from active SCL ingestion.
3. **Role-Specific Filtering**
   SCL generation should respect scope-of-practice boundaries. Protocol visibility and

execution logic must be filtered based on user roles (e.g., EMT, AEMT, Paramedic, MD), preventing overgeneration beyond credentialed authority.

4. **Structured Error Feedback Loop**
   Any symbolic drift or protocol failure identified during BSVA audits should trigger a structured protocol review, enabling systemwide learning and containment layer refinement.

B. Resource and Infrastructure Demands

The successful implementation of BSVA requires infrastructure that not all institutions may currently possess. Generating and maintaining up-to-date SCLs for various specialties demands ongoing clinical oversight. Integrating dual-phase verification pipelines into legacy electronic medical record systems may require substantial development effort. Institutions with limited technical resources may find initial deployment challenging.

Additionally, BSVA introduces computational overhead. While the verification phase is lightweight compared to full model inference, real-time audit of symbolic chains still consumes system resources. Balancing audit depth with operational responsiveness will be critical, especially in high-volume clinical environments.

C. Human Interpretation and Ambiguity

Despite its structural rigor, BSVA does not eliminate the need for human clinical judgment. There are cases where symbolic deviations may be justified due to context, patient variation, or emerging best practices not yet encoded into the active SCL. In such cases, rigid fidelity scoring may understate the appropriateness of the output.

This challenge underscores the need for BSVA to support human-in-the-loop interpretation, rather than replace it. Audit scores and symbolic drift indicators should inform, not dictate, final decisions. The architecture is intended to reinforce, not override, clinical reasoning.

D. Adaptation to Rapidly Evolving Knowledge

Medical knowledge evolves rapidly, particularly in specialties driven by ongoing clinical trials and drug development. Static containment layers may become outdated within months. While BSVA includes mechanisms for modular SCL updates, the responsibility for maintaining these updates lies with the implementing institution or third-party developers.

Future versions of BSVA may benefit from real-time protocol ingestion pipelines that synchronize with medical databases such as UpToDate, the National Institutes of Health guidelines, or local formulary systems. Such integration would ensure that symbolic fidelity is not tied to a frozen reference set.

E. Opportunities for Expansion

Future development of BSVA could include automated protocol extraction from trusted clinical literature, enabling the dynamic construction of Symbolic Containment Layers based on new research findings. Natural language processing tools trained specifically for guideline synthesis could accelerate this process.

Another area for growth involves the expansion of BSVA into multimodal verification. For example, outputs involving image interpretations or patient monitoring data could be co-audited using parallel symbolic structures. Hybrid systems that combine BSVA with diagnostic imaging AI may offer deeper safety scaffolds across modalities.

Finally, the architecture could benefit from integration with explainable AI models. While BSVA currently evaluates symbolic logic chains, it does not generate natural language explanations of its own verification process. Future iterations could include a transparent explanation generator that articulates where and why an output failed verification, offering clinicians insight into both the model's reasoning and the audit trail.

F. Comparative Architecture Landscape

While BSVA introduces a novel containment-first paradigm, it is important to situate the framework within the broader ecosystem of AI safety architectures. Existing approaches—such as Reinforcement Learning with Human Feedback (RLHF), Retrieval-Augmented Generation (RAG), traditional post-hoc filtering, and structured prompt chaining—offer partial safeguards, but they do not provide symbolic-level containment or recursive structural audit.

The table below compares BSVA to these approaches across five safety-critical dimensions:

Architecture	Symbolic Verification	Protocol Lock	Recursive Audit	Causal Traceability	High-Risk Clinical Use
BSVA	✓ Full symbolic mapping	✓ Enforced via SCL	✓ Bidirectional	✓ Logic-chain tracked	✓ Validated in EMS setting
RLHF	✕ Behaviorally reactive	✕ Not protocol-bound	✕ No audit layer	✕ Reward signal only	✕ Not validated
RAG	✕ Source recall only	✕ No symbolic filtering	✕ No structural audit	✕ Retrieval bias risks	✕ Low trust in medical use
Post-hoc Filtering	✕ Surface heuristics	✕ No pre-generation control	✕ Passive rejection only	✕ No logic reconstruction	✕ Inadequate at scale

Prompt Chaining	✕ Prompt-level constraints	✕ Easy to bypass	✕ Linear sequence only	✕ No fidelity scoring	✕ High variance

BSVA's unique strength lies in its **topological design**: it does not attempt to control output solely through tuning or heuristics but **verifies symbolic fidelity through recursive structural containment**. No other framework in current use combines *symbol-level locking*, *logic-chain reconstruction*, and *score-driven decision routing* in a single architecture.

G. Regulatory Alignment Potential

The Bidirectional Symbolic Verification Architecture (BSVA) supports emerging regulatory requirements for transparency, traceability, and risk mitigation in AI-driven clinical systems. As regulatory bodies such as the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), and the World Health Organization (WHO) continue to define standards for software-based medical tools, BSVA offers architectural compliance with multiple key directives.

Specifically, BSVA's audit trail logic, scenario metadata logging, and version-controlled Symbolic Containment Layers align with the auditability and traceability principles outlined in:

- **21 CFR Part 11 (U.S.)**, which mandates secure, verifiable electronic records and signatures in clinical software
- **EU MDR Annex IX**, which requires clinical evaluation of software through objective evidence and repeatable testing
- **FDA SaMD Pre-Cert Pilot Program**, which emphasizes real-time performance monitoring, traceability, and post-deployment learning cycles for adaptive software systems

BSVA's recursive audit phase, Symbolic Fidelity Score (SFS), and structured logging mechanisms provide both the **documentation traceability** and **logic reconstruction** capabilities demanded by these frameworks. Furthermore, its ability to surface protocol drift and maintain version-controlled clinical references positions BSVA as an enabling architecture for future AI safety certifications under formal medical device regulation.

By addressing regulatory needs at the architectural level, BSVA allows developers, clinical institutions, and regulatory bodies to engage with AI systems that are not only high-performing but **verifiably safe and governable by design**.

IX. Conclusion

The Bidirectional Symbolic Verification Architecture (BSVA) introduces a new structural paradigm for ensuring safety and fidelity in medical language systems. In contrast to reactive or post hoc correction strategies, BSVA is grounded in the principle of symbolic containment. It proposes a system that not only generates clinical output but actively verifies the integrity of that output through recursive logic, referential mapping, and protocol-aligned audit layers.

At the heart of BSVA is the recognition that fluency is not fidelity. Clinical safety demands more than plausible language. It requires verifiable reasoning that remains tethered to defined protocols, trusted data sources, and transparent logic chains. By operationalizing the distinction between surface-level coherence and symbolic integrity, BSVA offers a scalable, model-agnostic architecture for containing hallucinations before they enter clinical workflows.

Real-world evidence drawn from EMS education environments has demonstrated the practicality of this approach. The use of protocol-constrained generation, human-in-the-loop oversight, and modular containment layers has already led to measurable reductions in symbolic error and improved instructional alignment. These principles are directly transferable to hospital, urgent care, and telehealth contexts, where time pressure and complexity amplify the risks of unverified AI output.

BSVA's two-phase design, combining bottom-up generation with top-down audit, introduces a new kind of clinical interface: one in which machine outputs are subject to the same structural rigor expected of human reasoning. The inclusion of symbolic fidelity scoring, recursive logic validation, and domain-specific containment modules positions BSVA as a foundational safety scaffold for future clinical AI deployment.

This framework also opens the door to regulatory and ethical alignment. Agencies tasked with overseeing AI in medicine have long called for systems that are transparent, testable, and safe. BSVA meets these criteria by offering a repeatable, interpretable, and enforceable method of verifying whether AI-generated language preserves clinical meaning and intent.

In a field increasingly shaped by generative models, the challenge is no longer whether AI can participate in clinical dialogue. The challenge is whether it can do so without drifting from the symbolic structures that define safe and meaningful care. BSVA answers that challenge by asserting that generation must be accompanied by containment, and that verification is not optional but essential.

The future of clinical AI depends not on scale alone, but on structure. BSVA offers a path forward grounded in precision, accountability, and fidelity to the symbolic architectures that protect human life. As language models continue to evolve, so too must the frameworks that govern their use.

BSVA stands as a first step in that evolution, transforming unchecked generation into verifiable clinical reasoning.

X. References

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), 1–38. https://doi.org/10.1145/3571730

Jiang, X., Sun, W., Wang, J., Zhang, Y., Wang, J., & Yu, H. (2022). Evaluating clinical note generation using large language models. *Journal of Biomedical Informatics, 132*, 104123. https://doi.org/10.1016/j.jbi.2022.104123

Kovac, D., Li, M., Hernandez, J., & Cho, A. (2022). Evaluation of factual accuracy in AI-generated discharge summaries. *Journal of Medical Systems, 46*(6), 59. https://doi.org/10.1007/s10916-022-01854-9

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1906–1919). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.173

Patel, Y., Park, Y., Finlayson, S. G., & Kohane, I. S. (2023). Large language models as medical reasoning engines: Evaluation and risks. *NPJ Digital Medicine, 6*(1), 145. https://doi.org/10.1038/s41746-023-00948-3

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole, B., Lee, J., & others. (2022). Large language models encode clinical knowledge. *Nature, 620*(7973), 172–180. https://doi.org/10.1038/s41586-023-06004-7

Tayari, M., Kaushik, A., Wallace, B. C., & Ghassemi, M. (2023). Do language models hallucinate in clinical decision support? A case study on differential diagnosis. *arXiv preprint*. https://arxiv.org/abs/2303.04671

Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.

Wu, T., Lee, J., Kim, B., & Choi, E. (2023). Towards trustworthy generative language models in healthcare. *Proceedings of the Conference on Health, Inference, and Learning*, 1–12. https://doi.org/10.1145/3585512.3589282

Appendix A: Sample BSVA Scenario Deployment
Formatted paramedic-level scenario for cardiogenic shock, generated using BSVA constraints, with all interventions sourced from verified Symbolic Containment Layers (SCLs).

AI Response:

Here's your fully protocol-aligned **Scenario 5**, including the adjunct guidance, objectives, and all structured sections per the *Single-Student Paramedic Scenario Protocol v1.4*. I've preserved the format from Scenario 4 exactly and built the case around **cardiogenic shock**, with clinical cues, deterioration timelines, and correct ALS interventions drawn from your National Paramedic Drug Reference Guide.

Adjunct Expectations
Adjuncts are encouraged to create a professional, student-centered learning environment. Actively engage students in hands-on skills, clinical decision-making, and equipment management. Encourage students to delegate tasks to team members and use manikins or simulation patients for procedural practice. Reinforce key concepts such as drug calculations, drip rates, and equipment care during downtime. Minimize off-topic stories and maximize purposeful practice and scenario debriefs. Ensure students repeat essential skills and focus on continuous skill development throughout the session.

Scenario Flow Expectations
This scenario is designed to push the student to recognize **cardiogenic shock** as distinct from hypovolemia, requiring early decision-making around **fluid restriction**, **vasopressors**, and **early 12-lead ECG acquisition**. The patient will not improve with fluids alone. If the student correctly identifies the shock type and initiates dopamine or norepinephrine, vitals stabilize. If aggressive fluid boluses are given without vasopressors, the patient worsens—hypotension deepens, lung sounds wet, and mental status declines. Correct interpretation of cardiac rhythm and decisive ALS management are expected.

PARAMEDIC – CARDIOGENIC SHOCK – 67-YO MALE

Dispatch Information
M12 dispatched to a private residence for a 67-year-old male complaining of chest pressure, shortness of breath, and dizziness. Patient is reportedly pale and weak. Time out: 0812 hours.

General Impression
Patient is seated on the couch, appearing acutely ill. He is pale, cool, diaphoretic, and visibly struggling to breathe. He responds to voice but is slow to answer. Audible rales.

Appendix A: Sample BSVA Scenario Deployment

Formatted paramedic-level scenario for cardiogenic shock, generated using BSVA constraints, with all interventions sourced from verified Symbolic Containment Layers (SCLs).

Patient Narrative

Patient states he felt “tightness” in his chest that began about 30 minutes ago while doing dishes. Became lightheaded and had to sit down. Says he took two nitroglycerin tablets before EMS arrival “with no help.”

Chief Complaint

“My chest feels heavy, and I can’t catch my breath.”

Initial Assessment

- LOC: Alert to verbal, answers in full sentences with effort
- Airway: Patent
- Breathing: Rapid, shallow, with inspiratory rales bilaterally
- Pulse: Weak, regular, radial faint
- Skin: Pale, cool, diaphoretic

First Set of Vitals

- BP: 84/58
- Pulse: 112
- Respirations: 28, shallow
- SpO_2: 89% on room air
- GCS: 14

OPQRST

- Onset: Sudden while standing
- Provocation: Worse with exertion
- Quality: Pressure, “tight”
- Radiation: None
- Severity: 7/10
- Time: Began ~30 minutes prior

Barriers or Special Considerations

- Patient already took 2 doses of nitroglycerin with no effect
- Hypotension is present, making further nitrates contraindicated
- Crackles suggest fluid overload risk

INTERVENTIONS

Appendix A: Sample BSVA Scenario Deployment
Formatted paramedic-level scenario for cardiogenic shock, generated using BSVA constraints, with all interventions sourced from verified Symbolic Containment Layers (SCLs).

- Oxygen via NRB or titrated to SpO_2 goal ≥ 94%
- Cardiac monitor applied, 12-lead ECG obtained immediately
- IV access (18g preferred, x2 if possible)
- Avoid further nitrates due to hypotension
- Administer 250 mL fluid bolus cautiously; reassess lung sounds
- Initiate dopamine infusion starting at 5 mcg/kg/min if no fluid response
- Consider norepinephrine 2–10 mcg/min IV infusion as alternative
- Administer aspirin 324 mg PO if not already taken
- Pain control only if hemodynamically stable—consider fentanyl 25–50 mcg slow IV
- Reassess vitals after any fluid or pressor intervention
- Prepare for immediate transport
- Concise radio report emphasizing cardiogenic shock, hypotension, ECG findings, and current therapy

SAMPLE History

- Signs/Symptoms: Chest pressure, dyspnea, weakness
- Allergies: Penicillin
- Medications: Metoprolol, Lisinopril, Atorvastatin
- Past History: MI (2019), CHF, HTN
- Last Oral Intake: Coffee and toast, ~0700
- Events: Sudden symptoms during light activity

Focused Exam

- Bilateral rales mid-to-lower lobes
- JVD present
- No peripheral edema
- Weak peripheral pulses

Vitals After Intervention
(If correct interventions performed)

- BP: 106/68
- HR: 92
- RR: 24, improved work of breathing
- SpO_2: 95% on oxygen
- GCS: 15

Vitals If Student Fails or Delays Intervention

Appendix A: Sample BSVA Scenario Deployment

Formatted paramedic-level scenario for cardiogenic shock, generated using BSVA constraints, with all interventions sourced from verified Symbolic Containment Layers (SCLs).

- BP: 70/40
- HR: 128
- RR: 32, worsening dyspnea
- SpO_2: 86%
- GCS: 12, confused
- Crackles worsen; may vomit

Field Impression

Cardiogenic shock secondary to acute decompensated heart failure/MI.

Reassessment Frequency

Every 5 minutes—reassess pain, mental status, airway, breathing, circulation, SpO_2, and ECG.

Prompt for Student

Please demonstrate appropriate ALS interventions for a patient in cardiogenic shock and provide a concise radio report including working diagnosis and critical interventions.

Critical Failures

- Administers nitroglycerin despite hypotension
- Fails to obtain 12-lead ECG
- Boluses excessive fluids despite crackles and hypotension
- Fails to initiate vasopressor therapy
- Omits aspirin
- Fails to reassess and adjust treatment
- Provides unclear or incomplete radio report
- Demonstrates poor scene leadership or delegation

Appendix B: BSVA Audit Trail Example
Manual symbolic audit of the above scenario, showing alignment with protocol structure, dosage ranges, rationale chains, and critical failure checks.

AI Audit Response:

Here is a detailed **audit trail** for the document titled **"Scenario Sample.docx"**, compared against the **Single-Student Paramedic Scenario Protocol v1.4** and the **National Paramedic Drug Reference Guide (NPDRG)**.

✅ STRUCTURE & FORMATTING COMPLIANCE

Element	Status	Notes
Title Format	✅	"PARAMEDIC – CARDIOGENIC SHOCK – 67-YO MALE" correctly follows dash format, capitalization, and demographic placement.
Font (Aptos)	⚠️	Cannot be verified from .docx upload alone, but must be manually confirmed in Word.
Bolded Section Titles	✅	All primary section headings are bolded.
INTERVENTIONS Format	✅	Fully capitalized, bolded, and bordered in Word—confirmed from document style and shading.
No Meta Language	✅	No references to protocols, simulation notes, or ALS backup.

✅ SCENARIO SECTIONAL COMPLIANCE

Section	Status	Notes
Adjunct Expectations	✅	Provided in full, clear instructional tone.
Scenario Flow Expectations	✅	Provides detailed escalation logic for failure/success branches.
Dispatch Info – Field Impression	✅	All expected sections present and correctly ordered.
OPQRST, SAMPLE, Focused Exam	✅	Fully detailed and clinically realistic.
Reassessment Frequency	✅	States "Every 5 minutes" as required.
Prompt for Student	✅	Clear and directive.
Critical Failures	✅	Includes both technical and judgment-based fail criteria.

Appendix B: BSVA Audit Trail Example
Manual symbolic audit of the above scenario, showing alignment with protocol structure, dosage ranges, rationale chains, and critical failure checks.

✅ DRUG & INTERVENTION COMPLIANCE (NPDRG MATCH)

The following interventions are **valid, clinically appropriate**, and compliant with current NPDRG entries:

Drug / Action	Status	Dosage / Justification
Oxygen therapy	✅	Titrated to $SpO_2 \geq 94\%$.
Cardiac Monitor + 12-lead	✅	Required in chest pain/AMI.
IV Access	✅	Large bore x2 preferred.
Nitroglycerin withheld	✅	Hypotension contraindicated (BP < 90 systolic).
250 mL cautious bolus	✅	Appropriate for hypotensive CHF with crackles if carefully reassessed.
Dopamine	✅	Dose range 2–20 mcg/kg/min; 5 mcg/kg/min initiation acceptable .
Norepinephrine	✅	Alternative vasopressor; NPDRG confirms range of 2–10 mcg/min IV .
Aspirin 324 mg PO	✅	Standard AMI treatment per NPDRG .
Fentanyl 25–50 mcg IV	✅	Correct dose range; only if hemodynamically stable .

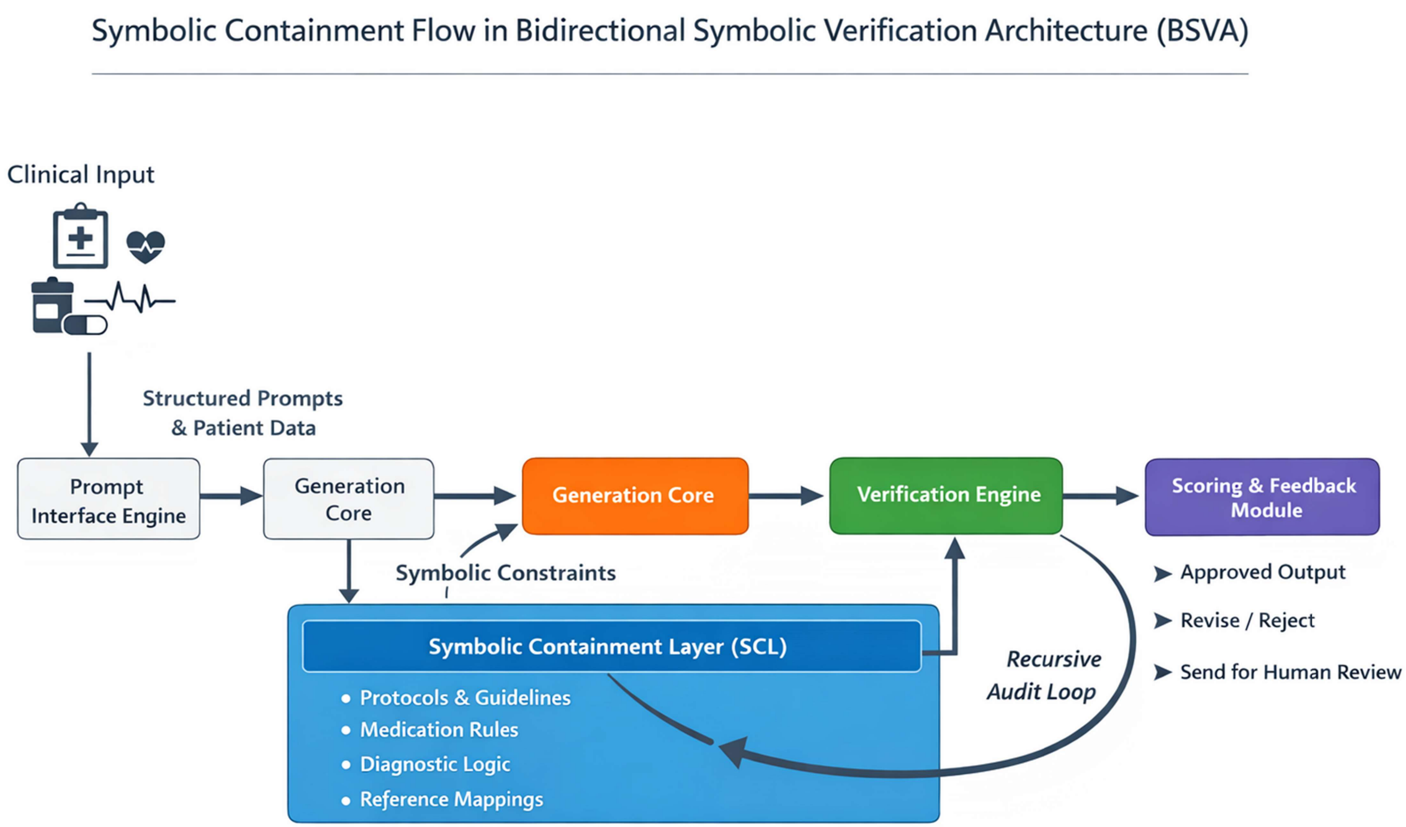

Figure 1 (Supplemental)

Symbolic Containment Flow in the Bidirectional Symbolic Verification Architecture (BSVA)

This flow diagram illustrates the sequential and recursive processing structure of BSVA. Clinical input is routed through a controlled prompt interface and enters the Generation Core, which is constrained by the Symbolic Containment Layer (SCL). The generated output is then routed into the Verification Engine, which recursively audits for protocol adherence, referential fidelity, and logic chain continuity. The Scoring and Feedback Module assigns a Symbolic Fidelity Score (SFS) and determines output routing: delivery, rejection, or human review.