TRACE WHITEPAPER

Title:
TRACE: A Protocol for Transparent, Verifiable, and Auditable AI Reasoning

Authors:
Carmen Kolcsar
SCALE UP PARTNERS S.R.L.
carmen@sup.partners

Abstract

Explainable artificial intelligence (XAI) is increasingly demanded in high-stakes applications, where transparency, verifiability, and accountability are critical. Existing methods often emphasize post-hoc interpretability or model-internal probing but fail to provide auditable, structured reasoning artifacts suitable for regulatory compliance and systematic human oversight. This paper introduces TRACE (Transparency, Reasoning, Assumptions, Confidence, Evidence), a protocol that externalizes AI reasoning as structured, machine-checkable artifacts designed specifically for compliance and audit use cases.

TRACE operationalizes NIST's Four Principles of Explainable AI and aligns with transparency mandates in the EU AI Act by providing formal specifications for reasoning documentation. We formalize the protocol algorithmically, present TRACE Lite as an efficiency-optimized variant for interactive use, and supply comprehensive JSON Schema specifications enabling automated validation and integration into AI assurance pipelines. The protocol bridges the gap between AI research and practical compliance requirements by producing reasoning artifacts that are simultaneously human-interpretable and machine-processable.

Illustrative case examples across law, healthcare, and education demonstrate domain-agnostic applicability, while detailed technical specifications ensure reproducibility and standardization. We provide complete implementation artifacts including prompt templates, validation schemas, and worked examples to facilitate adoption and further research. A comprehensive evaluation framework outlines methodologies for assessing TRACE effectiveness across technical performance, human factors, and compliance dimensions.

While empirical benchmarking is left for future work, this paper's contribution is conceptual, methodological, and infrastructural: TRACE enables AI systems to provide transparent, auditable, and domain-agnostic reasoning outputs that meet both human understanding needs and regulatory documentation requirements. By establishing standardized protocols for structured reasoning, this work positions TRACE as compliance-ready explainability infrastructure for trustworthy AI deployment.

Keywords: explainable AI, structured reasoning, AI compliance, transparency protocols, audit trails, regulatory alignment, AI governance

1. Introduction

AI-generated outputs are increasingly deployed in domains where decisions require traceability, contestability, and accountability—including law, healthcare, education, finance, and governance. As these applications scale from experimental tools to mission-critical systems, the gap between AI capabilities and explanation requirements becomes increasingly problematic. Current large language models (LLMs) typically produce fluent outputs with minimal insight into their reasoning processes, leaving users unable to audit assumptions, verify conclusions, or meet regulatory requirements for explainable decision-making.

The Explainability-Compliance Gap
Existing explainable AI (XAI) research has largely focused on either model-centric interpretability (understanding internal representations) or post-hoc rationalizations (natural language explanations generated after decisions). While valuable for research purposes, these approaches rarely provide the structured, auditable artifacts required for regulatory compliance or systematic human oversight. Legal frameworks like the EU AI Act Article 13 and governance standards like NIST AI RMF require explanations that are not just understandable, but formally documented, contestable, and integrated into organizational accountability systems.

Current Approaches Fall Short
Feature attribution methods (LIME, SHAP) excel at identifying important inputs but provide limited insight into reasoning processes. Chain-of-Thought prompting improves model performance but lacks the structured format and evidence citation needed for audit purposes. Constitutional AI incorporates reasoning principles but doesn't generate compliance-ready documentation. Model cards address system-level documentation but don't explain individual decisions. This landscape leaves practitioners with a fundamental choice between AI capabilities and regulatory requirements.

TRACE: Bridging Research and Compliance
We introduce TRACE (Transparency, Reasoning, Assumptions, Confidence, Evidence), a protocol designed specifically to address the explainability-compliance gap. Rather than attempting to interpret existing model outputs, TRACE structures the reasoning process itself, producing machine-readable artifacts that support both human understanding and automated compliance checking. TRACE operationalizes established XAI principles within a framework designed for practical deployment in regulated environments.

Key Innovation: Structured Reasoning as Compliance Infrastructure
TRACE's core innovation lies in treating explainable AI not as a post-processing step, but as a structured reasoning protocol that produces auditable artifacts by design. By formalizing reasoning chains, assumption statements, confidence assessments, and evidence citations within a standardized schema, TRACE enables AI systems to participate in formal

accountability processes while maintaining computational efficiency for real-world deployment.

AI-generated outputs are increasingly used in domains where decisions require **traceability and contestability**—law, healthcare, education, finance, and governance. Yet outputs from large language models (LLMs) typically lack reasoning transparency, leaving users unable to audit assumptions or assess reliability.

Most explainability research has focused on:

- **Model-centric interpretability** (e.g., saliency, attention analysis), or
- **Post-hoc rationalizations** (natural language explanations).

These approaches rarely provide structured artifacts that support **regulatory compliance** or **human oversight**.

We introduce **TRACE**, a domain-agnostic reasoning protocol designed to:

1. Externalize reasoning steps in a structured format.
2. Provide **explicit assumptions and confidence calibration**.
3. Cite supporting evidence for contestability.
4. Produce **machine-checkable outputs** (via JSON Schema).

Contributions:

- Formalization of the TRACE protocol.
- TRACE Lite, an efficiency-optimized variant.
- JSON Schema specifications for validation and compliance integration.
- Illustrative applications across law, healthcare, and education.
- Mapping to regulatory frameworks (EU AI Act, NIST AI RMF).

2. Related Work

2.1 Explainability Paradigms

The field of explainable AI (XAI) has evolved along several paradigms, each with distinct approaches to making AI decisions interpretable. Jacovi & Goldberg (2020) highlight the fundamental tension between faithful explanations (reflecting actual model internals) and plausible explanations (understandable but not necessarily faithful). This distinction is crucial for understanding where TRACE fits in the explainability landscape.

Lipton (2018) categorizes explainability approaches into transparency (understanding the model itself) and post-hoc interpretability (explaining specific decisions). Most current methods fall into the latter category, providing explanations after decisions are made rather than externalizing the reasoning process itself.

2.2 Model-Centric Interpretability

Traditional interpretability methods focus on understanding model internals through various techniques:

Feature Attribution Methods: LIME (Ribeiro et al., 2016) and SHAP (Lundberg & Lee, 2017) explain individual predictions by identifying important input features. While valuable for understanding which inputs matter, these methods provide limited insight into reasoning processes and lack structured formats for audit purposes.

Attention Analysis: In transformer-based models, attention weights are often interpreted as explanations (Jain & Wallace, 2019). However, attention-based explanations are unreliable indicators of model reasoning (Wiegreffe & Pinter, 2019).

Probing and Interpretation: Recent work attempts to understand what neural networks learn through probing classifiers and representation analysis (Rogers et al., 2020). These approaches are primarily research-oriented and don't provide actionable explanations for end users.

2.3 Structured Reasoning Approaches

Several recent approaches attempt to structure AI reasoning more explicitly:

- Chain-of-Thought (CoT) Prompting: Wei et al. (2022) demonstrate that prompting large language models to show their reasoning steps improves performance on complex tasks. However, CoT is designed for performance enhancement rather than auditability, lacks structured formats, and provides no evidence citation mechanisms.
- Tree-of-Thought: Yao et al. (2023) extend CoT by exploring multiple reasoning paths simultaneously. While sophisticated, this approach focuses on problem-solving efficiency rather than creating auditable reasoning artifacts.
- Constitutional AI: Anthropic's constitutional AI approach (Bai et al., 2022) incorporates explicit principles into AI reasoning. However, it doesn't provide structured outputs suitable for compliance documentation or systematic audit.
- Reasoning Frameworks: Several domain-specific reasoning frameworks exist, such as legal reasoning systems (Ashley, 2017) and medical diagnostic aids (Shortliffe & Buchanan, 1975), but these are typically domain-specific and lack standardized output formats.

2.4 Compliance-Oriented Explainability

Growing regulatory pressure has led to frameworks specifically designed for compliance and documentation:

- Model Cards: Mitchell et al. (2019) introduced model cards for documenting ML model characteristics, intended uses, and limitations. While valuable for model-level documentation, model cards don't address instance-level decision explanations.
- Algorithmic Impact Assessments: Reisman et al. (2018) propose systematic evaluation of algorithmic systems' societal impacts. These frameworks focus on system-level assessment rather than individual decision explanations.

- GDPR Article 22 Compliance: Wachter et al. (2017) analyze explanation requirements under EU data protection law, highlighting the need for meaningful explanations that are contestable and actionable. Current technical approaches often fail to meet these legal standards.
- Regulatory Frameworks: Recent policy documents, including NIST AI RMF (2023) and the EU AI Act (2024), establish principles for trustworthy AI but lack specific technical implementation guidance for explanation systems.

2.5 Comparison with Existing Approaches

Approach	Reasoning Structure	Auditability	Evidence Citation	Standards Alignment	Machine Readable	Real-time Feasible
LIME/SHAP	None	Low	No	None	Partial	Yes
Chain-of-Thought	Informal	Low	No	None	No	Yes
Constitutional AI	Principle-based	Medium	No	Partial	No	Yes
Model Cards	Static documentation	High	Yes	High	Yes	No
Algorithmic IA	System-level	High	Yes	High	Yes	No
TRACE	Formal protocol	High	Yes	NIST/EU	Yes	Yes

2.6 Gap Analysis

Despite significant progress in XAI research, several critical gaps remain for high-stakes applications:

1. Structured Output Formats: Most explanation methods produce natural language or visualizations that are difficult to automatically process, validate, or integrate into compliance workflows.
2. Evidence Linkage: Few approaches provide systematic citation of supporting evidence, making it difficult to verify or contest explanations.
3. Confidence Calibration: While some methods provide confidence scores, these are rarely well-calibrated or structured for decision-making.
4. Regulatory Alignment: Technical explainability methods are often developed independently of regulatory requirements, creating gaps between what researchers build and what practitioners need.
5. Auditability at Scale: Most current approaches require manual review of each explanation, making large-scale audit processes impractical.

TRACE addresses these gaps by providing a structured, evidence-linked, confidence-calibrated protocol specifically designed for compliance and audit use cases while remaining computationally feasible for real-time applications.

3. The TRACE Protocol

3.1 Formal Definition

TRACE defines a reasoning artifact as:

$\mathcal{E}(Q) = \langle T, R, A, C, Ev \rangle$

Where:

- **Q (Query):** The user input query
- **T (Transparency):** Problem restatement and decomposition.
- **R (Reasoning):** Stepwise premises, logic, conclusions, confidence.
- **A (Assumptions):** Contextual boundaries and limitations.
- **C (Confidence):** Overall certainty, qualitative (🟢, 🟡, 🔴) and optionally quantitative.
- **E (Evidence):** Cited references supporting claims.

3.2 Algorithmic Specification

Algorithm 1: TRACE Protocol

```
Input:

    User query Q

Output:

    Structured reasoning artifact ℰ(Q) = ⟨T, R, A, C, E⟩

where:

    T  denotes the transparency component,

    R  denotes the reasoning component,

    A  denotes the assumptions component,

    C  denotes the global confidence assessment, and

    E denotes the evidence component.

Procedure:

    1. Construct T by restating Q in plain language and decomposing it into
its key components.
```

2. Identify the relevant knowledge, applicable frameworks, and known limitations associated with Q.

3. Construct R by iterating over each inference step and recording:

    (a) the premise,

    (b) the applied logic,

    (c) the resulting conclusion, and

    (d) a step-level confidence value in {high, medium, low}.

4. Construct A by listing the contextual assumptions and knowledge boundaries that influence the reasoning process.

5. Generate a synthesis by integrating the reasoning steps into a coherent overall answer.

6. Construct C by assigning an overall confidence value in {high, medium, low}, together with a justification.

7. Construct E by attaching references, sources, or supporting materials for the principal claims.

8. Perform a validation check by identifying fragile assumptions, potential reasoning errors, and missing or uncertain evidence.

9. Return the structured reasoning artifact $\mathcal{E}(Q) = \langle T, R, A, C, E \rangle$.

4. TRACE Lite

4.1 Motivation

Full TRACE improves transparency but increases token usage, latency, and cost. TRACE Lite offers a compact format for interactive or large-scale use.

4.2 Protocol Specification

```
## Q
Restate the question in 1–2 sentences.

## R
Reasoning steps in bullets (max 4).

## A
Key assumptions (2–3 bullets).

## C
Conclusion (1–2 sentences) + confidence 🟢🟡🔴.

## L
Limitations (1–2 bullets).
```

TRACE Lite retains **transparency, reasoning, assumptions, confidence, and limitations**, with ~70% fewer tokens.

4.5 Evaluation Framework

While empirical validation of TRACE is beyond the scope of this specification paper, establishing clear evaluation criteria is essential for future research and practical adoption. We propose a multi-dimensional evaluation framework that addresses technical performance, human factors, and compliance effectiveness.

4.5.1 Technical Performance Metrics

Completeness Assessment:

- **Coverage Ratio**: Proportion of reasoning steps explicitly documented vs. human expert baseline
- **Component Completeness**: Percentage of TRACE artifacts (T,R,A,C,E) successfully generated
- **Schema Compliance**: Rate of valid JSON outputs meeting specification requirements

Consistency Evaluation:

- **Inter-run Reliability**: Agreement between multiple TRACE executions on identical inputs
- **Cross-model Consistency**: Alignment of TRACE outputs across different LLM implementations
- **Temporal Stability**: Consistency of explanations over time for equivalent queries

Efficiency Metrics:

- **Latency Overhead**: Additional response time compared to unexplained outputs
- **Token Efficiency**: TRACE vs. TRACE Lite token usage comparison
- **Computational Cost**: Resource consumption for generating structured reasoning

4.5.2 Human Factors Assessment

Comprehensibility Studies:

- **Expert Evaluation**: Domain experts' ability to understand and validate TRACE reasoning chains
- **Novice Accessibility**: Non-experts' comprehension of TRACE outputs in their simplified form
- **Cross-domain Transferability**: Understanding consistency across legal, medical, and educational contexts

Auditability Measurement:

- **Verification Time**: Duration required for domain experts to audit TRACE reasoning chains
- **Error Detection Rate**: Ability of human reviewers to identify flawed reasoning using TRACE artifacts
- **Evidence Validation Efficiency**: Time to verify cited sources and their relevance

Trust Calibration:

- **Appropriate Reliance**: Whether TRACE improves users' ability to rely appropriately on AI outputs
- **Confidence Alignment**: Correlation between TRACE confidence indicators and actual accuracy
- **Uncertainty Communication**: Effectiveness of limitation statements in conveying reasoning boundaries

4.5.3 Compliance Assessment

Regulatory Acceptability:

- **Legal Expert Review**: Assessment by regulatory compliance professionals
- **Audit Trail Completeness**: Coverage of requirements for contestable decision-making
- **Documentation Standards**: Alignment with industry-specific compliance frameworks

Operational Integration:

- **Workflow Compatibility**: Ease of integrating TRACE into existing audit and review processes
- **Scalability for Compliance**: Feasibility of using TRACE for large-scale regulatory reporting
- **Cost-Benefit Analysis**: Compliance value vs. implementation overhead

4.5.4 Proposed Validation Studies

Study 1: Cross-Domain Expert Evaluation

- **Participants**: 15 experts each in law, healthcare, and education (N=45)
- **Design**: Within-subjects comparison of TRACE vs. standard AI outputs
- **Measures**: Comprehensibility ratings, verification time, error detection accuracy
- **Duration**: 6 months for recruitment, data collection, and analysis

Study 2: Regulatory Compliance Assessment

- **Participants**: Compliance officers, auditors, and regulatory affairs professionals
- **Design**: Evaluation of TRACE outputs against specific regulatory standards (EU AI Act Article 13, NIST AI RMF)

- **Measures**: Compliance coverage ratings, audit efficiency improvements, documentation quality
- **Duration**: 8 months including regulatory expert recruitment

Study 3: Longitudinal Adoption Study

- **Setting**: Partner organizations implementing TRACE in production systems
- **Design**: Before-after comparison of audit processes and decision quality
- **Measures**: Audit time reduction, error rates, user satisfaction, regulatory review outcomes
- **Duration**: 12 months for meaningful organizational change assessment

4.5.5 Benchmark Development

TRACE Benchmark Suite: We propose developing a standardized benchmark consisting of:

- **Reasoning Quality Dataset**: 1000 queries across domains with expert-validated reasoning chains
- **Compliance Test Cases**: Scenarios specifically designed to test regulatory alignment
- **Edge Case Collection**: Queries that challenge TRACE protocol limits and assumptions
- **Cross-cultural Validation**: Multi-language and cultural context testing

Evaluation Metrics Standard:

- **Reasoning Fidelity Score**: Semantic similarity between TRACE reasoning and expert-generated reasoning
- **Audit Efficiency Index**: Ratio of audit time with TRACE vs. without TRACE
- **Compliance Coverage Metric**: Percentage of regulatory requirements addressed by TRACE artifacts

4.5.6 Comparison Methodology

Future empirical work should compare TRACE against existing approaches using:

Baseline Methods:

- Unstructured natural language explanations
- Chain-of-Thought prompting outputs
- LIME/SHAP feature attribution
- Domain-specific explanation systems

Comparison Dimensions:

- Time to understand and verify explanations
- Accuracy of human decisions based on explanations
- User confidence calibration
- Compliance documentation completeness

Statistical Framework:

- Power analysis for detecting meaningful differences in audit efficiency
- Multilevel modeling accounting for domain, user expertise, and task complexity
- Bayesian approaches for incorporating prior knowledge about explanation effectiveness

5. Governance Mapping

EU AI Act Alignment

- **Transparency (Art. 13):** T, R, and E fields directly support this.
- **Reliability & Contestability:** Confidence and Validation steps.
  ⚠️ **Gap:** System-level dataset lineage and audit logs not covered.

NIST AI RMF Alignment

- **Map:** Problem decomposition.
- **Measure:** Confidence calibration and assumptions.
- **Manage:** Validation step.
  ⚠️ **Gap:** Bias measurement and lifecycle monitoring remain external.

6. Illustrative Applications

- **Law:** Applying precedents in contract disputes.
- **Healthcare:** Diagnostic reasoning under missing lab data.
- **Education:** Solving math word problems with explicit reasoning steps.

These examples show TRACE's **domain-agnostic adaptability**.

6.1 Detailed Case Study: Legal Contract Analysis

To illustrate TRACE protocol implementation, we present a complete example from legal practice: analyzing whether a software licensing agreement contains adequate data protection clauses for EU AI Act compliance.

Query: "Does this software licensing agreement provide adequate data protection provisions for deploying AI systems under EU AI Act requirements?"

Full TRACE Output:

```
json

{
```

"version": "1.0.0",

"task_id": "legal-001-gdpr-ai-act",

"query": "Does this software licensing agreement provide adequate data protection provisions for deploying AI systems under EU AI Act requirements?",

"t": {

"restatement": "Evaluate whether the licensing agreement's data protection clauses meet EU AI Act requirements for high-risk AI system deployment, focusing on data governance, human oversight, and liability allocation.",

"components": [

"Data processing governance clauses",

"Liability allocation for AI system failures",

"Human oversight and intervention requirements",

"Compliance documentation obligations",

"Cross-border data transfer provisions"

]

},

"r": [

{

"premise": "EU AI Act Article 10 requires high-risk AI systems to have appropriate data governance measures",

"logic": "License must allocate responsibility for data quality, bias monitoring, and dataset documentation between licensor and licensee",

"conclusion": "Agreement lacks specific AI data governance allocation - contains only generic GDPR compliance clauses",

"confidence_step": "high",

"claims": [

"EU AI Act Article 10 mandates data governance for high-risk AI", "Current agreement contains only GDPR-standard data protection clauses",

"AI-specific data governance requirements exceed standard GDPR provisions"

]

},

{

"premise": "Article 14 requires human oversight capabilities and clear responsibility chains",

"logic": "License should specify which party maintains human oversight controls and incident response procedures",

"conclusion": "Agreement includes general liability terms but no AI-specific human oversight responsibilities",

"confidence_step": "medium",

"claims": [

"Human oversight is mandatory under Article 14",

"Agreement assigns general software liability but not AI oversight duties",

"Unclear division of responsibility for AI system monitoring"

]

},

{

"premise": "Article 16 establishes accuracy and robustness requirements with ongoing monitoring obligations",

"logic": "License must clarify who bears responsibility for post-deployment performance monitoring and corrective measures",

"conclusion": "No provisions for ongoing AI system performance monitoring or accuracy validation responsibilities",

"confidence_step": "high",

"claims": [

"Article 16 requires ongoing accuracy monitoring",

"Agreement lacks post-deployment monitoring clauses",

"No mechanism for addressing performance degradation"
]
}
],

"a": [
"Assumes software qualifies as 'high-risk AI system' under Annex III classifications",
"Assumes EU jurisdiction applies to deployment context",
"Analysis limited to explicitly stated contract terms, not implied obligations",
"Does not consider potential regulatory updates or interpretive guidance"
],

"c": {
"label": "medium",
"justification": "Clear gaps identified in AI-specific governance provisions, but standard GDPR clauses provide some baseline protection. Confidence limited by evolving regulatory interpretation.",
"probability": 0.75,
"calibration_method": "heuristic"
},

"e": [
{
"id": "eu-ai-act-2024",
"kind": "report",
"title": "Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act)",

```
      "year": 2024,

      "locator": "https://eur-lex.europa.eu/legal-
content/EN/TXT/?uri=CELEX:32024R1689",

      "accessed_at": "2024-12-15T10:00:00Z",

      "supports_claims": [0, 1, 2]

    },

    {

      "id": "gdpr-2016",

      "kind": "article",

      "title": "General Data Protection Regulation (GDPR)",

      "year": 2016,

      "locator": "https://eur-lex.europa.eu/eli/reg/2016/679/oj",

      "accessed_at": "2024-12-15T10:00:00Z",

      "supports_claims": [0]

    }

  ],



  "synthesis": "The licensing agreement provides standard GDPR-compliant
data protection but lacks AI Act-specific provisions for high-risk systems.
Key gaps include AI data governance allocation, human oversight
responsibilities, and ongoing performance monitoring obligations. Agreement
requires AI Act-specific amendments for compliant deployment.",



  "validation": {

    "potential_errors": [

      "May have missed implied obligations under general contract law",

      "AI Act interpretation still evolving - guidance documents may alter
requirements",

      "Jurisdictional assumptions may not apply to specific deployment
context"

    ],
```

```
    "fragile_assumptions": [
      "Classification as 'high-risk' AI system",
      "EU jurisdiction application",
      "No additional contractual side letters or amendments"
    ],
    "evidence_gaps": [
      "No access to AI Act implementation guidelines",
      "Missing industry-specific compliance interpretations",
      "No precedent cases for contractual AI Act compliance"
    ]
  }
}
```

TRACE Lite Version for Same Query:

```json
{
  "version": "1.0.0-lite",
  "task_id": "legal-001-lite",
  "query": "Does this software licensing agreement provide adequate data protection provisions for deploying AI systems under EU AI Act requirements?",

  "q": "Assess whether license meets EU AI Act data protection requirements for high-risk AI systems, focusing on governance and oversight provisions.",

  "r": [
    "EU AI Act Article 10 requires specific data governance beyond GDPR - agreement has only standard GDPR clauses",
    "Article 14 mandates human oversight responsibilities - license lacks AI-specific oversight allocation",
```

```
    "Article 16 requires ongoing monitoring obligations - no post-deployment performance provisions",

    "Standard liability terms insufficient for AI Act compliance requirements"

  ],



  "a": [

    "Assumes high-risk AI system classification under Annex III",

    "EU jurisdiction applies to deployment",

    "Analysis based only on explicit contract terms"

  ],



  "c": {

    "label": "medium",

    "text": "Clear gaps in AI-specific provisions, but baseline GDPR protection exists. Requires AI Act amendments for compliance."

  },


  "l": [

    "AI Act interpretation still evolving, implementation guidance pending",

    "Analysis limited to contract text without industry-specific context"

  ]

}
```


Analysis of Example:
This worked example demonstrates how TRACE structures complex legal reasoning while maintaining auditability. The full version provides comprehensive documentation suitable for compliance review, while TRACE Lite offers the same core reasoning in a format suitable for quick consultation. Both versions maintain evidence linkage and confidence assessment while acknowledging analytical limitations.

7. Discussion

TRACE = **structured explainability protocol**.
TRACE Lite = **efficiency-optimized profile**.
Together, they provide a **protocol family** balancing auditability and scalability.

8. Limitations

- TRACE explanations may be **plausible, not faithful** to model internals.
- No empirical benchmarks in this paper.
- Governance gaps: dataset lineage, system logs, bias metrics not addressed.

8.1 Fundamental Limitations

Plausibility vs. Faithfulness Trade-off
TRACE explanations represent plausible reasoning chains that may not faithfully reflect the actual computational processes within large language models. This limitation is inherent to post-hoc explanation methods and cannot be fully resolved without fundamental changes to model architectures. TRACE prioritizes auditability and regulatory compliance over perfect faithfulness, following the pragmatic approach that useful, contestable explanations are preferable to opaque but "faithful" systems.

Reasoning Quality Dependency
The quality of TRACE outputs is fundamentally limited by the underlying model's reasoning capabilities. Models with poor domain knowledge or logical reasoning abilities will produce structured but potentially flawed TRACE artifacts. This limitation is particularly concerning in high-stakes domains where reasoning errors could have serious consequences.

Human Oversight Requirement
TRACE does not eliminate the need for human expert review; it structures and facilitates such review. The protocol assumes availability of domain experts capable of validating reasoning chains, which may not be feasible in all deployment contexts or at all scales.

8.2 Technical Constraints

Computational Overhead
Full TRACE protocol significantly increases computational requirements compared to standard AI outputs. Our preliminary estimates suggest 40-70% increases in token usage and 25-50% increases in latency, depending on query complexity. While TRACE Lite mitigates these costs, it does so at the expense of reasoning depth and evidence detail.

Schema Rigidity
The JSON schema structure, while enabling machine validation, may not capture all forms of valid reasoning. Complex arguments involving analogical reasoning, implicit cultural knowledge, or highly domain-specific logical patterns may be forced into inappropriate structural templates, potentially losing nuance or introducing artifacts.

Evidence Source Limitations
TRACE's evidence citation capability is constrained by the underlying model's training data and knowledge cutoffs. The protocol cannot cite sources that weren't present during training or provide real-time access to updated information without external integration.

Confidence Calibration Challenges
Current confidence indicators in TRACE (🟢🟡🔴) are based on heuristic assessments rather than rigorous probability calibration. Without empirical validation studies, these confidence levels may not accurately reflect true likelihood of correctness, potentially misleading users about reasoning reliability.

8.3 Scalability and Deployment Challenges

Organizational Integration Complexity
Implementing TRACE requires significant changes to existing AI deployment pipelines, audit processes, and compliance workflows. Organizations must train staff to interpret TRACE outputs, modify approval processes to accommodate structured reasoning review, and potentially redesign user interfaces to display complex reasoning artifacts effectively.

Cross-Domain Validation Requirements
While TRACE is designed to be domain-agnostic, its effectiveness likely varies significantly across application areas. Legal reasoning, medical diagnosis, and educational assessment may require domain-specific adaptations that haven't been systematically explored or validated.

Regulatory Evolution
Current alignment with NIST AI RMF and EU AI Act requirements may become outdated as regulations evolve. TRACE's static schema design may require frequent updates to maintain compliance relevance, potentially creating versioning and backward compatibility challenges.

8.4 Governance and Bias Concerns

Bias Amplification Risk
Structured reasoning in TRACE format may make certain types of bias more visible while potentially obscuring others. The protocol doesn't include built-in bias detection mechanisms, and the requirement for explicit reasoning steps might inadvertently legitimize biased reasoning by giving it formal structure.

System-Level Audit Gaps
TRACE addresses instance-level explanation but doesn't cover system-level concerns such as training data lineage, model selection processes, or deployment monitoring. Integration with broader AI governance frameworks remains an open challenge.

Cultural and Linguistic Limitations
The current TRACE specification reflects Western, English-language reasoning patterns and legal frameworks. Adaptation to different cultural contexts, legal systems, and languages may require fundamental protocol modifications rather than simple translation.

8.5 Research Gaps and Future Work Priorities

Immediate Research Priorities (0-6 months)

1. **Empirical Validation Study**: Controlled evaluation of TRACE effectiveness across domains
   - Cross-domain expert evaluation with healthcare, legal, and educational professionals
   - Comparison with existing explainability methods on standardized tasks
   - User study measuring comprehension, trust calibration, and audit efficiency
2. **Confidence Calibration Research**: Development of principled approaches to uncertainty quantification
   - Integration with formal uncertainty quantification methods
   - Validation of qualitative confidence indicators against ground truth
   - Cross-model calibration consistency studies

Short-term Development (6-18 months)

3. **Domain-Specific Adaptations**: Specialized TRACE variants for high-stakes applications
   - Legal TRACE: Integration with case law citation and precedent analysis
   - Clinical TRACE: Alignment with medical reasoning frameworks and evidence-based medicine
   - Educational TRACE: Pedagogical reasoning patterns and learning objective alignment
4. **Scalability Engineering**: Technical infrastructure for large-scale deployment
   - Distributed validation systems for JSON schema checking
   - Integration APIs for existing AI governance platforms
   - Performance optimization for real-time applications
5. **Bias Detection Integration**: Systematic approaches to identifying and flagging problematic reasoning
   - Automated detection of common reasoning fallacies
   - Integration with fairness assessment tools
   - Cultural sensitivity validation frameworks

Long-term Research Directions (18+ months)

6. **Faithful Explanation Research**: Exploration of architectures that could provide both structure and faithfulness
   - Integration with interpretable-by-design model architectures
   - Causal reasoning integration for more robust explanations
   - Multi-step verification systems for reasoning chain validation
7. **Regulatory Co-evolution**: Dynamic adaptation to evolving compliance requirements
   - Automated regulatory requirement extraction and protocol updates
   - Cross-jurisdictional compliance mapping
   - Integration with emerging AI governance standards (ISO/IEC 42001, IEEE standards)
8. **Human-AI Collaboration Enhancement**: Optimizing TRACE for human-AI team decision-making
   - Adaptive explanation detail based on user expertise
   - Interactive reasoning chain refinement

- Integration with human decision-support workflows

8.6 Known Risks and Mitigation Strategies

Risk: Over-reliance on Structured Format
Users may trust TRACE outputs more than warranted simply because they appear systematic and well-documented.
Mitigation: Include prominent limitation statements and encourage critical evaluation training.

Risk: Gaming and Manipulation
Sophisticated users might learn to generate TRACE-compliant outputs that appear reasonable but contain subtle flaws.
Mitigation: Develop adversarial testing frameworks and red-team evaluation protocols.

Risk: Regulatory Compliance Theater
Organizations might use TRACE for compliance purposes without a genuine commitment to explainable AI principles.
Mitigation: Emphasize that TRACE is a tool for genuine transparency, not mere documentation generation.

9. Future Work

- Empirical benchmarking (calibration, verification speed, trust).
- Human evaluation studies.
- Integration with assurance pipelines (ISO/IEC 42001).
- Domain-specific adaptations (legal TRACE, clinical TRACE).

10. Conclusion

TRACE introduces a standardized reasoning protocol for transparent, verifiable, and auditable AI reasoning, directly addressing the growing gap between AI capabilities and regulatory requirements for explainable systems. Unlike existing XAI approaches that focus primarily on model interpretability or post-hoc rationalization, TRACE provides a structured framework for externalizing reasoning processes in machine-readable formats designed for compliance and audit workflows.

Technical Contribution

TRACE's formal specification, including algorithmic definitions and comprehensive JSON schemas, establishes a foundation for standardized reasoning documentation across AI applications. The protocol's modular design—encompassing transparency, reasoning chains, assumptions, confidence assessment, and evidence citation—provides comprehensive coverage of explainability requirements while maintaining computational feasibility through the TRACE Lite efficiency optimization.

Infrastructure for Research Community

By publishing complete technical specifications, validation schemas, and implementation examples, this work provides immediately usable infrastructure for both researchers and practitioners. The protocol's domain-agnostic design enables cross-disciplinary adoption while supporting domain-specific adaptations through its flexible schema architecture. Future empirical research can build upon these foundations to validate TRACE effectiveness and develop domain-specific refinements.

Compliance and Governance Impact

TRACE's explicit alignment with NIST AI Risk Management Framework and EU AI Act requirements positions it as a practical solution for organizations facing regulatory compliance challenges. The protocol's structured output format enables automated compliance checking and systematic audit processes, potentially reducing the administrative burden of AI governance while improving transparency quality.

Immediate Adoption Potential

The combination of formal specification, efficiency optimization, and comprehensive documentation enables immediate integration into existing AI systems. Organizations can begin implementing TRACE protocols without waiting for extensive validation studies, using the framework to improve their AI explainability practices while contributing to broader empirical evaluation efforts.

Future Research Directions

The evaluation framework outlined in this paper provides a roadmap for systematic validation across technical performance, human factors, and compliance dimensions. Priority areas include confidence calibration research, cross-domain validation studies, and integration with emerging AI governance standards. The protocol's modular design facilitates incremental improvements and domain-specific adaptations as research progresses.

TRACE represents a convergence of AI research and practical governance needs, providing structured reasoning capabilities that advance both scientific understanding and regulatory compliance. By establishing standardized protocols for AI explainability, this work contributes essential infrastructure for the responsible deployment of AI systems in high-stakes applications where transparency, accountability, and human oversight are paramount.

References

Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age.* Cambridge University Press.

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., ... & Kaplan, J. (2022). *Constitutional AI: Harmlessness from AI feedback*. arXiv preprint arXiv:2212.08073.

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union, L 1689, 1-144.

Jacovi, A., & Goldberg, Y. (2020). *Towards faithfully interpretable NLP systems: How should we define and evaluate faithfulness?* In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 4198-4205).

Jain, S., & Wallace, B. C. (2019). *Attention is not explanation.* In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 3543-3556).

Lipton, Z. C. (2018). *The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.* Queue, 16(3), 31-57.

Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions.* In Advances in Neural Information Processing Systems (pp. 4765-4774).

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019). *Model cards for model reporting*. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220-229).

NIST. (2023). *AI Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Phillips, P. J., Hahn, C. A., Fontana, P. C., Yates, A. N., Greene, K., Broniatowski, D. A., & Przybocki, M. A. (2021). *Four principles of explainable artificial intelligence (NIST IR 8312).* National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8312

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). *Algorithmic impact assessments: A practical framework for public agency accountability.* AI Now Institute.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why should I trust you?" Explaining the predictions of any classifier.* In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). *A primer in BERTology: What we know about how BERT works.* Transactions of the Association for Computational Linguistics, 8, 842-866.

Shortliffe, E. H., & Buchanan, B. G. (1975). *A model of inexact reasoning in medicine.* Mathematical Biosciences, 23(3-4), 351-379.

Turpin, M., Michael, J., Perez, E., & Bowman, S. R. (2023). *Language models don't always say what they think: Unfaithful explanations in chain-of-thought reasoning.* In Advances in Neural Information Processing Systems (Vol. 36, pp. 74952-74965).

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). *Why a right to explanation of automated decision-making does not exist in the general data protection regulation.* International Data Privacy Law, 7(2), 76-99.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). *Chain-of-thought prompting elicits reasoning in large language models.* In Advances in Neural Information Processing Systems (Vol. 35, pp. 24824-24837).

Wiegreffe, S., & Pinter, Y. (2019). *Attention is not not explanation.* In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (pp. 11-20).

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). *Tree of thoughts: Deliberate problem solving with large language models. In Advances in Neural Information Processing Systems (Vol. 36, pp. 11809-11822).*

Phillips, P. J., Hahn, C. A., Fontana, P. C., Yates, A. N., Greene, K., Broniatowski, D. A., & Przybocki, M. A. (2021). *Four principles of explainable artificial intelligence*. NIST IR 8312.

Jacovi, A., & Goldberg, Y. (2020). *Towards faithfully interpretable NLP models*. ACL.

Turpin, M., et al. (2023). *Language models don't always say what they think: Unfaithful explanations in chain-of-thought reasoning*. NeurIPS.

Wei, J., et al. (2022). *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. NeurIPS.

Appendix A: Full TRACE Prompt Specification (Verbose Version)

The following is the **complete prompt template** used in TRACE testing. It is published here for reproducibility.

```
<reasoning_protocol>

STEP 1: PROBLEM DECOMPOSITION
- Restate the question in plain words
- Identify key components
- List what must be determined

STEP 2: KNOWLEDGE RETRIEVAL
- What information is known?
- What frameworks or principles apply?
- What are knowledge limits?

STEP 3: REASONING CHAIN
For each step:
```

```
- Premise: [state fact/rule]
- Logic: [explain connection]
- Conclusion: [state outcome]
- Confidence: [🟢/🟡/🔴 + justification]

STEP 4: SYNTHESIS
- Connect pieces
- Overall conclusion
- Alternatives considered

STEP 5: VALIDATION
- What could make this wrong?
- Fragile assumptions
- Missing evidence

CONFIDENCE SCALE
🟢 High (90%+)
🟡 Medium (60-89%)
🔴 Low (<60%)

OUTPUT FORMAT:
## ANALYSIS
[Decomposition]
## REASONING CHAIN
[Step-by-step logic]
## CONCLUSION
[Final answer + confidence]
## LIMITATIONS
[Missing pieces]

</reasoning_protocol>
```

Appendix B: JSON Schema Specifications

B.1 TRACE Core Schema

```
{
  "$id": "https://example.org/schemas/trace-core.schema.json",
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "title": "TRACE Core Reasoning Artifact",
  "description": "Minimal, domain-agnostic schema for TRACE outputs.",
  "type": "object",
  "required": ["version", "task_id", "query", "t", "r", "a", "c", "e"],
  "properties": {
    "version": {
      "type": "string",
      "const": "1.0.0",
      "description": "TRACE protocol spec version."
    },
    "task_id": {
      "type": "string",
      "minLength": 1,
      "description": "Caller-supplied identifier for the task/instance."
    },
    "query": {
      "type": "string",
```

```
      "minLength": 1,
      "description": "Original user query Q."
    },

    "t": {
      "title": "Transparency",
      "type": "object",
      "required": ["restatement", "components"],
      "properties": {
        "restatement": { "type": "string", "minLength": 1, "maxLength":
2000 },
        "components": {
          "type": "array",
          "minItems": 1,
          "items": { "type": "string", "minLength": 1, "maxLength": 500 }
        }
      }
    },

    "r": {
      "title": "Reasoning",
      "type": "array",
      "minItems": 1,
      "items": {
        "type": "object",
        "required": ["premise", "logic", "conclusion", "confidence_step"],
        "properties": {
          "premise": { "type": "string", "minLength": 1, "maxLength": 2000
},
          "logic":   { "type": "string", "minLength": 1, "maxLength": 2000
},
          "conclusion": { "type": "string", "minLength": 1, "maxLength":
2000 },
          "confidence_step": {
            "type": "string",
            "enum": ["high", "medium", "low"]
          },
          "claims": {
            "description": "Atomic propositions introduced in this step for
evidence checking.",
            "type": "array",
            "items": { "type": "string", "minLength": 1, "maxLength": 1000
}
          }
        }
      }
    },

    "a": {
      "title": "Assumptions",
      "type": "array",
      "minItems": 1,
      "items": { "type": "string", "minLength": 1, "maxLength": 1000 }
    },

    "c": {
      "title": "Confidence (overall)",
      "type": "object",
      "required": ["label"],
```

```
        "properties": {
          "label": { "type": "string", "enum": ["high", "medium", "low"] },
          "justification": { "type": "string", "minLength": 1, "maxLength":
2000 },
          "probability": {
            "type": "number",
            "minimum": 0.0,
            "maximum": 1.0,
            "description": "Optional numeric calibration if available."
          },
          "calibration_method": {
            "type": "string",
            "enum": ["none", "heuristic", "isotonic", "temperature",
"other"],
            "default": "none"
          }
        }
      },

      "e": {
        "title": "Evidence",
        "type": "array",
        "minItems": 1,
        "items": {
          "type": "object",
          "required": ["id", "kind"],
          "properties": {
            "id": { "type": "string", "minLength": 1, "description": "Local
citation key (e.g., smith2021)." },
            "kind": { "type": "string", "enum": ["url", "doi", "book",
"article", "report", "dataset", "other"] },
            "title": { "type": "string" },
            "authors": { "type": "array", "items": { "type": "string" } },
            "year": { "type": "integer", "minimum": 1800, "maximum": 2100 },
            "locator": {
              "type": "string",
              "description": "URL, DOI, or repository path",
              "pattern": "^(10\\.\\d{4,9}/[-._;()/:A-Z0-9]+|https?://.+|[A-
Za-z0-9._/:-]+)$",
              "examples": ["10.6028/NIST.IR.8312",
"https://example.org/doc.pdf"]
            },
            "accessed_at": { "type": "string", "format": "date-time" },
            "supports_claims": {
              "type": "array",
              "description": "Indices of `r[*].claims` this evidence supports
(optional).",
              "items": { "type": "integer", "minimum": 0 }
            }
          }
        }
      },

      "synthesis": {
        "type": "string",
        "description": "Overall integration summary.",
        "maxLength": 3000
      },
```

```
    "validation": {
      "type": "object",
      "required": ["potential_errors", "fragile_assumptions",
"evidence_gaps"],
      "properties": {
        "potential_errors": {
          "type": "array", "minItems": 1,
          "items": { "type": "string", "minLength": 1, "maxLength": 1000 }
        },
        "fragile_assumptions": {
          "type": "array", "minItems": 1,
          "items": { "type": "string", "minLength": 1, "maxLength": 1000 }
        },
        "evidence_gaps": {
          "type": "array", "minItems": 1,
          "items": { "type": "string", "minLength": 1, "maxLength": 1000 }
        }
      }
    }
  },
  "additionalProperties": false
}
```

B.2 TRACE Extended Schema (with governance metadata)

```
{
  "$id": "https://example.org/schemas/trace-extended.schema.json",
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "title": "TRACE Extended Reasoning Artifact",
  "allOf": [
    { "$ref": "https://example.org/schemas/trace-core.schema.json" },
    {
      "type": "object",
      "properties": {
        "meta": {
          "type": "object",
          "required": ["created_at", "locale", "generator"],
          "properties": {
            "created_at": { "type": "string", "format": "date-time" },
            "locale": { "type": "string", "pattern": "^[a-z]{2}(-[A-
Z]{2})?$", "examples": ["en", "en-US", "ro-RO"] },
            "generator": {
              "type": "object",
              "required": ["model_name"],
              "properties": {
                "model_name": { "type": "string" },
                "model_version": { "type": "string" },
                "decoding": {
                  "type": "object",
                  "properties": {
                    "temperature": { "type": "number", "minimum": 0 },
                    "top_p": { "type": "number", "minimum": 0, "maximum": 1
},

                    "max_new_tokens": { "type": "integer", "minimum": 1 }
                  },
                  "additionalProperties": false
                }
```

```
              },
              "additionalProperties": false
            },
            "context_tokens": { "type": "integer", "minimum": 0 },
            "output_tokens": { "type": "integer", "minimum": 0 }
          },
          "additionalProperties": false
        },
        "provenance": {
          "type": "object",
          "description": "Optional system-level hooks for governance
alignment.",
          "properties": {
            "dataset_version": { "type": "string" },
            "policy_profile": { "type": "string", "description": "e.g.,
NIST RMF profile id or internal control set." },
            "trace_id": { "type": "string", "description": "Link key to org
logging pipeline." },
            "hash": {
              "type": "string",
              "description": "Integrity hash over canonicalized JSON of
this artifact.",
              "pattern": "^[A-Fa-f0-9]{64}$"
            },
            "signature": { "type": "string", "description": "Detached
signature (e.g., JWS) for non-repudiation." }
          },
          "additionalProperties": false
        }
      },
      "required": ["meta"]
    }
  ]
}
```

B.3 TRACE Lite Profile

```
{
  "$id": "https://example.org/schemas/trace-lite.schema.json",
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "title": "TRACE Lite Profile",
  "description": "Efficiency-optimized profile: Q/R/A/C/L with token
budgets.",
  "type": "object",
  "required": ["version", "task_id", "query", "q", "r", "a", "c", "l"],
  "properties": {
    "version": { "type": "string", "const": "1.0.0-lite" },
    "task_id": { "type": "string", "minLength": 1 },
    "query": { "type": "string", "minLength": 1 },

    "q": { "type": "string", "maxLength": 400, "description": "Restatement
(1-2 sentences)." },

    "r": {
      "type": "array", "minItems": 1, "maxItems": 4,
      "items": { "type": "string", "minLength": 1, "maxLength": 300 }
    },
```

```
    "a": {
      "type": "array", "minItems": 1, "maxItems": 3,
      "items": { "type": "string", "minLength": 1, "maxLength": 200 }
    },

    "c": {
      "type": "object",
      "required": ["label"],
      "properties": {
        "label": { "type": "string", "enum": ["high", "medium", "low"] },
        "text": { "type": "string", "maxLength": 300 }
      }
    },

    "l": {
      "type": "array", "minItems": 1, "maxItems": 2,
      "items": { "type": "string", "minLength": 1, "maxLength": 200 }
    }
  },
  "additionalProperties": false
}
```


B.4 Example Instances

- Healthcare case (Core schema instance)


```
{
  "version": "1.0.0",
  "task_id": "case-001",
  "query": "Is intervention X clinically advisable for symptom cluster
Y?",
  "t": {
    "restatement": "Assess whether X is appropriate for Y given limited
lab data.",
    "components": ["evidence for X→Y", "contraindications",
"uncertainty"]
  },
  "r": [
    {
      "premise": "Guideline G recommends X for moderate Y",
      "logic": "Patient symptoms map to moderate severity per G",
      "conclusion": "X is indicated",
      "confidence_step": "medium",
      "claims": ["G supports X for moderate Y", "Patient meets moderate
criteria"]
    },
    {
      "premise": "Missing lab values",
      "logic": "Contraindications may be undetected",
      "conclusion": "Delay until labs or use conservative dose",
      "confidence_step": "low",
      "claims": ["Labs absent could hide risks"]
    }
  ],
  "a": ["Assume no drug interactions", "Assume adherence to G applies"],
  "c": {
    "label": "medium",
    "justification": "Guideline alignment but missing labs",
```

```
      "probability": 0.7,
      "calibration_method": "heuristic"
    },
    "e": [
      {
        "id": "guidelineG2022",
        "kind": "article",
        "title": "Guideline G for Y",
        "year": 2022,
        "locator": "https://example.org/guidelineG",
        "accessed_at": "2025-08-28T10:00:00Z",
        "supports_claims": [0]
      }
    ],
    "synthesis": "X likely appropriate with caution pending labs.",
    "validation": {
      "potential_errors": ["Guideline misapplied to atypical
presentation"],
      "fragile_assumptions": ["No interactions"],
      "evidence_gaps": ["Recent RCTs on X for comorbidity Z"]
    }
}
```

- Math problem (Lite schema instance)

```
{
  "version": "1.0.0-lite",
  "task_id": "math-017",
  "query": "Solve: If a:b=2:3 and b:c=4:5, find a:c",
  "q": "Combine the two ratios to express a in terms of c.",
  "r": [
    "a:b=2:3 ⇒ a=2k, b=3k",
    "b:c=4:5 ⇒ b=4m, c=5m",
    "Match b: 3k=4m ⇒ k=4t, m=3t",
    "Then a=8t, c=15t ⇒ a:c=8:15"
  ],
  "a": ["Positive real ratios", "No hidden constraints"],
  "c": { "label": "high", "text": "Direct proportion manipulation" },
  "l": ["Assumes exact ratios; measurement noise ignored"]
}
```