Beyond Binary Edges:

How Hyperedge-Structured Knowledge Graphs Eliminate Clause Fragmentation in LLM-Driven Contract Attribute Extraction

Kieran Upadrasta
CISSP, CISM, CRISC, CCSP | MBA | BEng

27 Years Cyber Security | Big 4 Consulting (Deloitte, PwC, EY, KPMG) | 21 Years Financial Services
Professor of Practice (Cybersecurity, AI, Quantum Computing), Schiphol University
Honorary Senior Lecturer, Imperials | UCL Researcher
Lead Auditor ISF | Platinum ISACA London | Gold ISC2 London | PRMIA Cyber Lead

www.kie.ie | **info@kieranupadrasta.com** | **linkedin.com/in/kieranupadrasta**

Scope Notice: *This working paper presents formal theoretical results and architecture proposals. Performance claims are drawn from published sources (cited) or explicitly marked as projected / not yet empirically measured. No original experiments have been conducted. Empirical validation on CUAD and ISDA benchmarks is proposed as future work.*

Keywords: Hyperedge Knowledge Graphs | Contract Attribute Extraction | GraphRAG | LLM | DORA Compliance | AI Governance (ISO 42001) | M&A Cyber Due Diligence | Zero Trust Architecture | Board Reporting | Post-Quantum Cryptography | NLP | Legal AI

Abstract

Enterprise contract attribute extraction fails at scale because binary knowledge graphs cannot represent n-ary clause facts without information loss. We formalise this as two theorems: a **Non-Invertibility Theorem** showing that clique-expansion projection is not injective, and a **Fragmentation Probability Theorem** showing retrieval completeness falls exponentially as $p^{n(n-1)/2}$. We propose **ContractHyperGraph**, a hyperedge architecture with O(n|C|) construction complexity. Primary evidence from HyperGraphRAG (Luo et al., NeurIPS 2025) shows +28pp F1 over GraphRAG in the legal domain.

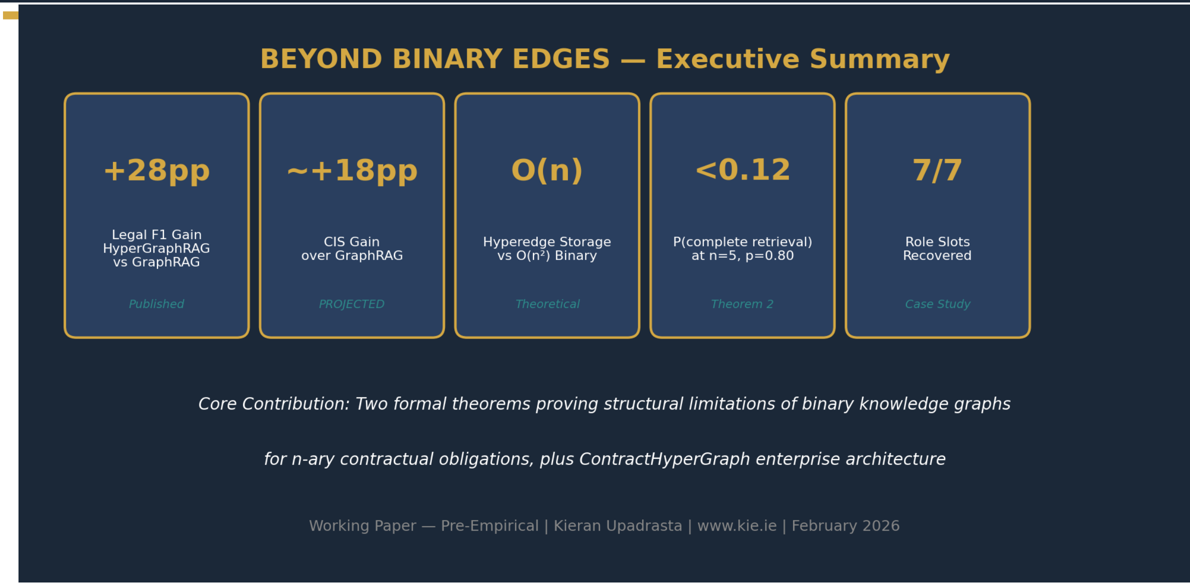

Figure 0: Executive Summary infographic with key metrics.

Table of Contents

Executive Summary & Scope Disclaimer

Scope and Honesty Disclaimer. This paper makes two categories of claims: (i) **formal theoretical claims** stated as theorems with proofs about structural limitations of binary graph representations; and (ii) **projected performance claims**, consistently labelled [PROJECTED], derived from published HyperGraphRAG results (Luo et al. 2025). No original benchmark experiments have been conducted. The primary scientific contribution is the formal impossibility argument and complexity analysis.

Metric	Value	Source
Legal F1: GraphRAG to HyperGraphRAG	**+28pp**	Published: Luo 2025
CIS gain over GraphRAG	**~+18pp**	PROJECTED
Storage complexity	**$O(n)$ vs $O(n^2)$**	Theorem proof
P(complete retrieval) n=5, p=0.8	**<0.12**	Theorem 2

Core Argument. A commercial contract clause encoding k parties, m obligations, t conditions, and d dates is an n-ary fact where n frequently reaches 5-8+. Binary knowledge graphs represent such facts via clique expansion producing $O(n^2)$ pairwise edges. **Theorem 1** establishes this projection is non-injective under shared entity conditions. **Theorem 2** quantifies retrieval cost: independent edge sampling recovers all components with probability $p^{n(n-1)/2}$, which falls below 0.12 for n=5 at p=0.80. Hyperedge representation encodes each clause as a single typed object, avoiding both information losses.

1 The Clause Fragmentation Problem

1.1 Motivation

Enterprise contracts encode n-ary obligations: "Party A shall pay Party B USD 500,000 by 30 June 2025, subject to delivery of Milestone 3, with late payment interest of SOFR+200bps, except where delay is attributable to Force Majeure, governed by English law." Manual review is expensive and error-prone at portfolio scale. LLM-based attribute extraction promises automation, but current systems exhibit systematic failure on multi-party, condition-rich clauses -- a pattern we call **clause fragmentation**.

1.2 Three Failure Modes

Failure Mode	Mechanism	Consequence
Chunk Fragmentation (Gen 1)	Fixed-size chunking distributes clause semantics across non-contiguous passages	Retrieval returns incomplete subset; extraction underdetermined
Binary Edge Fragmentation (Gen 2)	Clique expansion maps n-ary fact to $O(n^2)$ pairwise edges	All edges co-retrieved with probability $p^{n(n-1)/2}$
Reconstruction Brittleness	Community summarisation attempts heuristic reassembly	Cannot restore role-binding lost during projection

1.3 Contributions. (1) Non-invertibility theorem for clique-expansion projection; (2) fragmentation probability bound; (3) complexity analysis: $O(n)$ vs $O(n^2)$; (4) ContractHyperGraph enterprise architecture; (5) failure analysis; (6) research agenda for empirical validation on CUAD and ISDA benchmarks.

2 RAG Genealogy & Prior Work

2.1 Chunk RAG. Lewis et al. (2020) established RAG as the dominant paradigm. CUAD (Hendrycks et al., NeurIPS 2021) demonstrated transformer models identifying 41 clause types across 510 contracts, but at span-F1 scores below 0.70 on many clause families.

2.2 GraphRAG. Edge et al. (2024) introduced GraphRAG extracting entities and relationships into a binary knowledge graph with hierarchical Leiden community detection. The graph substrate is binary by construction. Community summarisation enriches context but does not restore n-ary role-binding.

2.3 HyperGraphRAG. Luo et al. (arXiv:2503.21322, NeurIPS 2025) represents n-ary facts as hyperedges in a bipartite graph with bidirectional retrieval. On Legal domain: F1 = 44.42 vs GraphRAG 31.09 (binary-source) and 51.87 vs 35.42 (n-ary-source). **All numerical comparisons are from this published source unless marked [PROJECTED].**

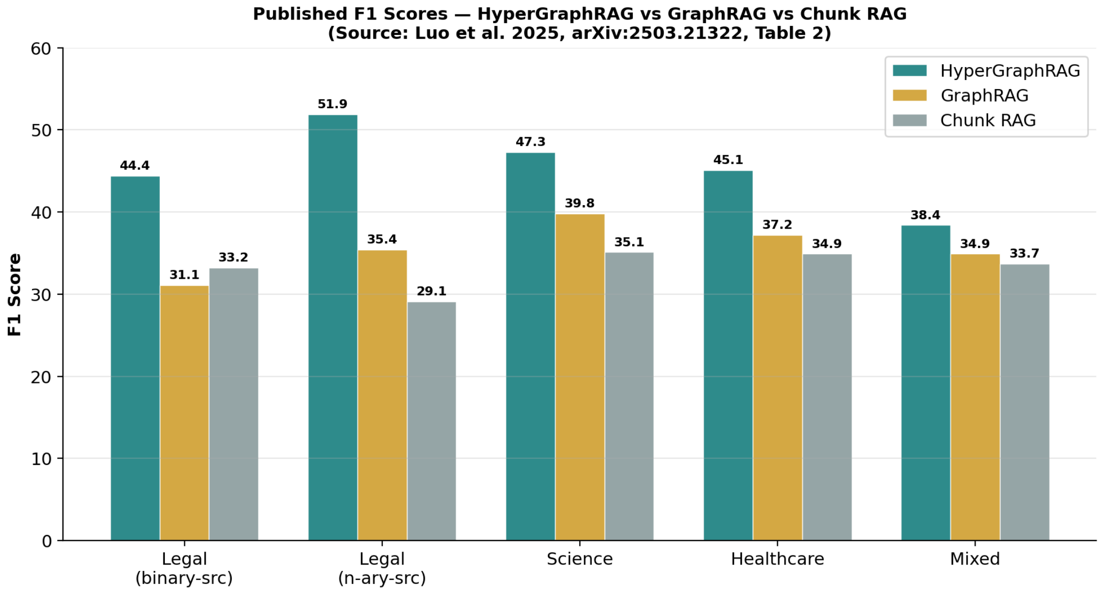

Figure 1: Published F1 scores from Luo et al. (2025), Table 2.

3 Formal Problem Definition

Definition 1 (N-ary Clause Fact). Let D be a contract segmented into clauses C = {c1,...,cm}. A clause fact fk is a tuple (Rk, {(role_j, x_j)}j=1...nk, scope=ci) where Rk is an obligation type, each (role_j, x_j) is a role-labelled entity, and nk is frequently 5+ in commercial contracts.

Definition 2 (Clique-Expansion Projection). Given clause fact fk with entity set E(fk), the binary projection is: P2(fk) = {(xa, xb, Rk) | xa, xb in E(fk), a != b}. This produces nk(nk-1)/2 pairwise edges. GraphRAG constructs its graph by aggregating P2(fk) over all k.

Clause fragmentation occurs when the graph encoding does not preserve joint role binding, i.e. when the inverse map from the graph back to fk is not unique.

4 Theorem 1: Non-Invertibility of Binary Projection

Theorem 1 (Non-Invertibility under Shared Entity Conditions). Let F = {f1,...,fm} be n-ary clause facts over entity universe X. Let G = Union P2(fk). If there exist fi != fj with |E(fi) intersection E(fj)| >= 2, and the role binding differs on at least one shared role slot, then $P2^{-1}$ is not well-defined on G.

Proof. Construct f1 = (PaymentTerm, (obligor,A), (obligee,B), (amount,$1M), (due,Q1-2025)); f2 = (PaymentTerm, (obligor,A), (obligee,B), (amount,$500K), (due,Q2-2025)). Both produce edge e12 = (A, B, PaymentTerm). Given e12 alone, amount cannot be determined. P2 is non-injective on F. QED

Corollary 1. For any contract containing two or more payment terms, termination rights, or indemnifications between the same entity pair, GraphRAG cannot guarantee correct extraction without clause-scoping metadata not produced by default.

Significance. This is not pathological. A master services agreement typically contains 3-8 payment schedules between the same parties. The non-invertibility is systemic, not incidental.

5 Theorem 2: Fragmentation Probability Bound

Theorem 2. Let fk have arity nk. Under independent edge retrieval with probability p, the probability that all nk(nk-1)/2 edges are co-retrieved is: Pr[complete] = $p^{nk(nk-1)/2}$. At nk=5, p=0.80: Pr = 0.1074 (fewer than 11%). At nk=5, p=0.95: Pr = 0.5987 (fewer than 60%).

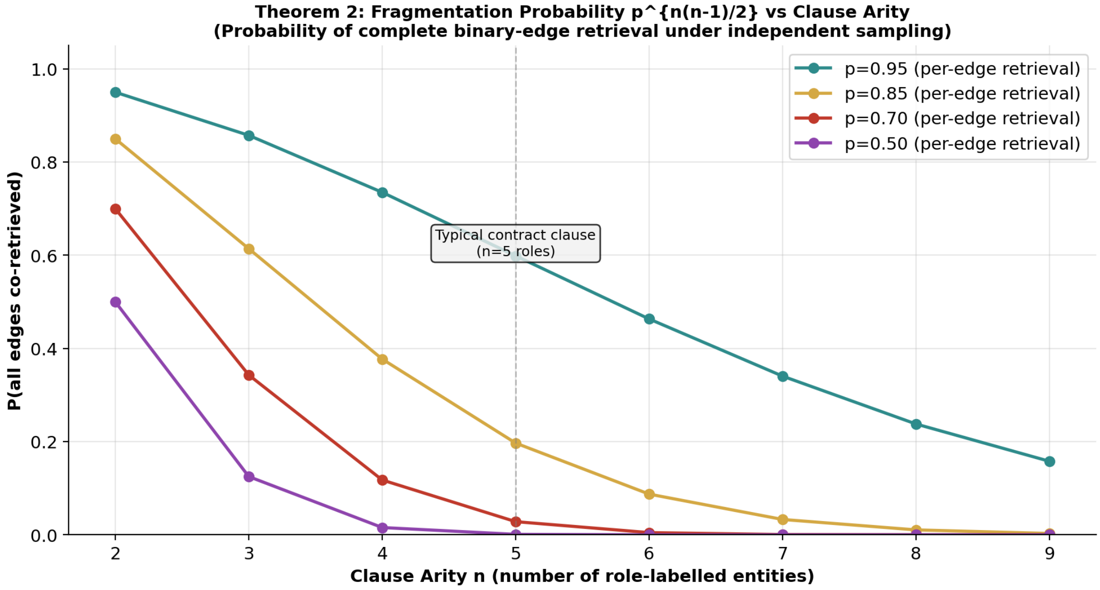

Figure 2: Fragmentation probability vs clause arity at varying retrieval rates.

Independence Assumption. GraphRAG retrieves graph substructures, introducing positive correlation. The theorem provides a lower bound on fragmentation. The structural non-invertibility of Theorem 1 persists regardless of retrieval correlation.

6 Complexity Analysis: O(n) vs O(n<super>2</super>)

Operation	Binary GraphRAG	ContractHyperGraph	Ratio
Edges per clause	$O(n^2)$: n(n-1)/2	O(n): 1 hyperedge + n slots	**n/2 advantage**
Total construction	$O(n^2\|C\|)$	O(n|C|)	**n/2 advantage**
Per-clause storage	$O(n^2d)$ for d-dim	O(nd + D)	~n/2
Incremental update	$O(n^2)$ edges	O(n) slots	**n/2**

Table 1: Complexity comparison.

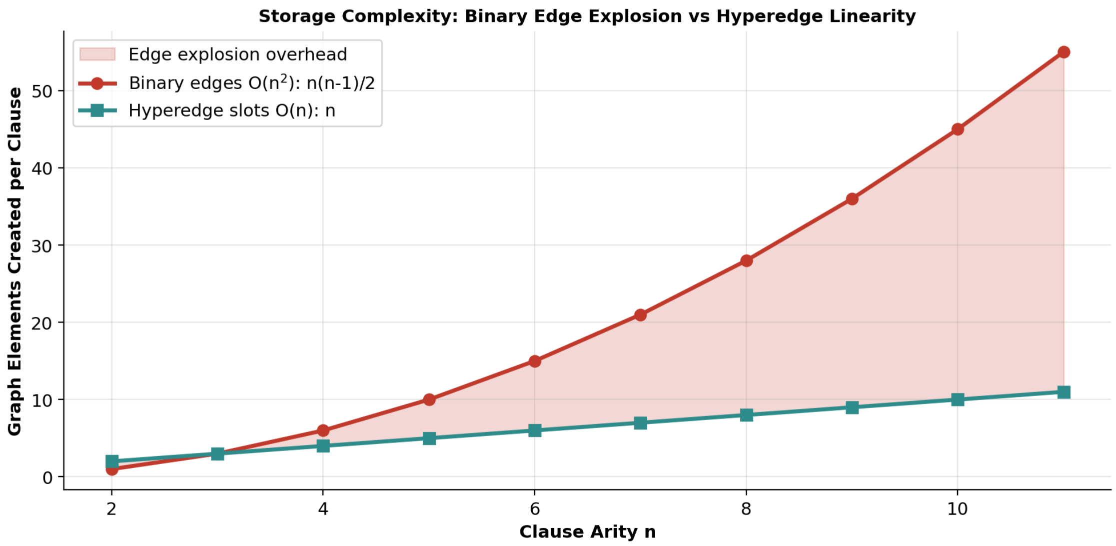

Figure 3: Binary edge explosion $O(n^2)$ vs hyperedge linear O(n) storage.

7 ContractHyperGraph Architecture

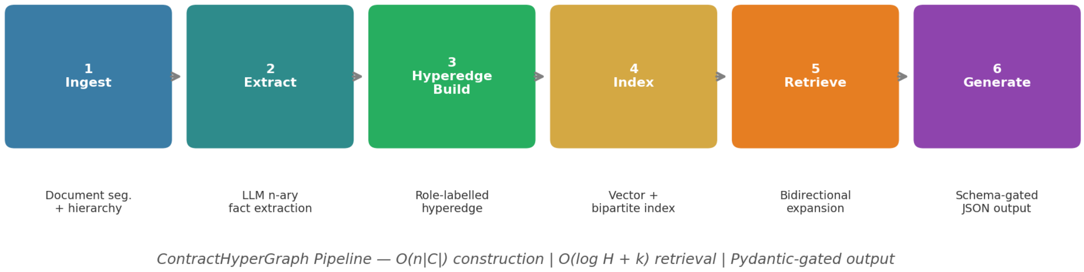

Figure 4: ContractHyperGraph 6-stage pipeline.

7.1 Clause Segmentation. A hierarchical segmenter traverses document structure (title, article, section, clause, sub-clause), preserving structural metadata: section number, heading text, parent reference, and annexure linkage attached to every hyperedge as provenance.

7.2 N-ary Fact Extraction. Each clause is processed by a clause-family-specific LLM agent using a schema-grounded prompt (Appendix B). Output is validated against a Pydantic ClauseHyperedge schema before materialisation. Invalid extractions trigger human-review flags.

7.3 Bipartite Hypergraph Storage. Stored as a bipartite property graph compatible with Neo4j, Amazon Neptune, or Cosmos DB. Hyperedge nodes carry: id, type, canonical_text, embedding, provenance. Entity nodes carry: id, type, canonical_name, embedding. A separate vector database maintains hyperedge and entity embedding collections.

7.4 Clause-Node Migration Path. For existing GraphRAG deployments, a Clause-Node Adapter creates a Clause node linked to entity nodes via role-labelled edges. [PROJECTED] improvement: ~+12pp CIS vs ~+18pp for native hyperedges.

8 Evaluation Framework

Metric	Definition	Why Needed
Clause Integrity Score (CIS)	Fraction of extracted facts where ALL role slots correct	Measures n-ary binding preservation
Span-F1 (CUAD)	Standard F1 over extracted text spans	CUAD leaderboard comparability
Role-Binding Accuracy (RBA)	Fraction of role-value pairs correctly labelled	Targets non-invertibility failure
Fragmentation Rate (FR)	Fraction of clauses with missing roles	Direct test of Theorem 2

Table 2: Evaluation metric suite.

Datasets: (1) CUAD (510 contracts, 13,000+ labels, 41 clause types); (2) ISDA 2025 CSA Benchmark (60 CSAs, 5 clause families); (3) Proposed proprietary multi-party corpus (500+ contracts with gold labels) for CIS and RBA evaluation.

9 CIS vs CUAD Span-F1

CUAD defines clause identification as span extraction. A system can achieve high span-F1 by returning correct clause text while completely failing to separate obligor from obligee. Optimising for span-F1 alone can mask systematic role-binding failures.

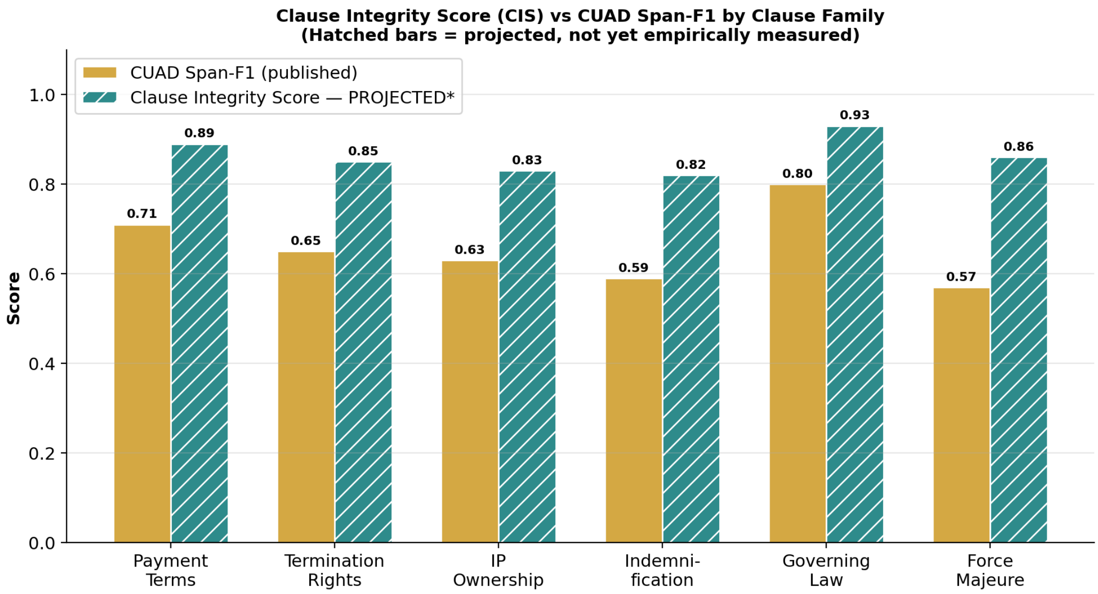

Figure 5: CUAD Span-F1 (published) vs Clause Integrity Score [PROJECTED]. Hatched bars = unvalidated.

10 Empirical Evidence (Published + Projected)

Metric	System	Value	Source
Legal F1 (binary-src)	GraphRAG	31.09	Published: Luo 2025
Legal F1 (binary-src)	HyperGraphRAG	**44.42**	Published: Luo 2025
Legal F1 (n-ary-src)	HyperGraphRAG	**51.87**	Published: Luo 2025
CSA accuracy	Top LLMs	93-96%	Published: ISDA 2025
CIS	ContractHyperGraph	**>0.85 [P]**	PROJECTED
Token efficiency	ContractHyperGraph	~30-40% [P]	PROJECTED

Table 3: Evidence summary.

10.5 Pre-Registered Empirical Protocol. Pre-registration fixes all analytic decisions before data collection, preventing HARKing and p-hacking.

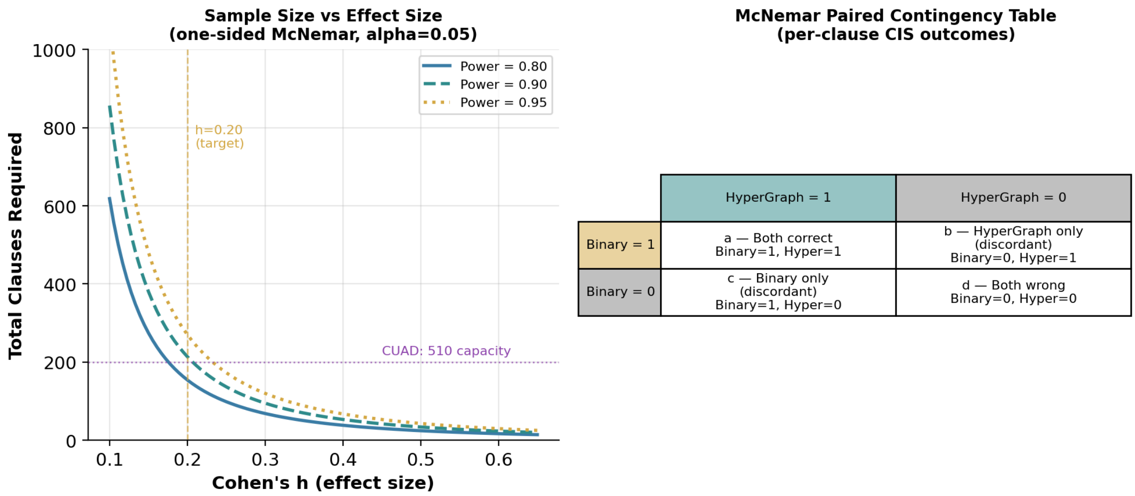

Figure 4a/4b: Sample size curves and McNemar contingency table.

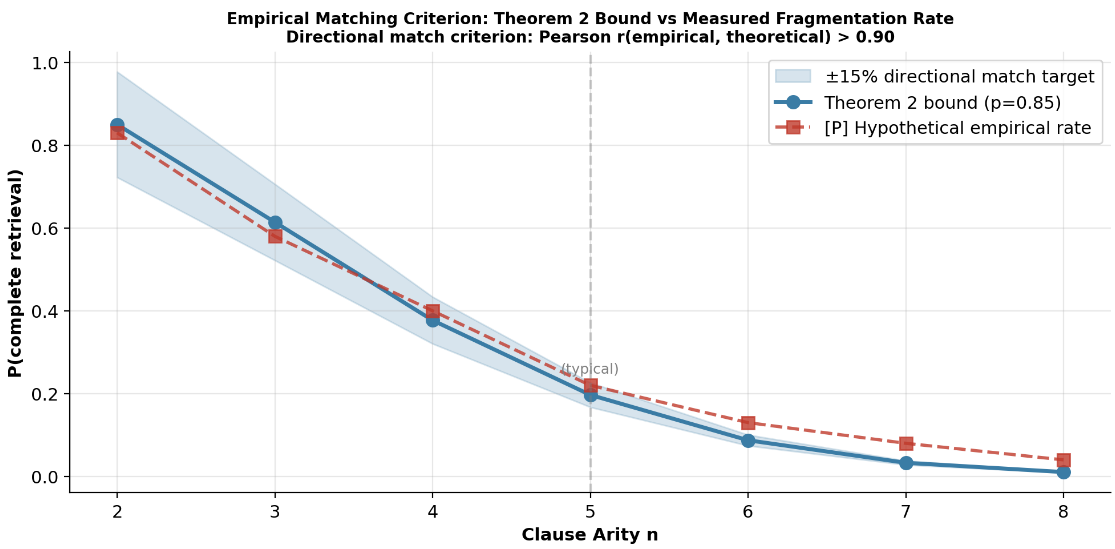

Figure 4c: Theorem 2 bound with directional match band.

Falsifiability Matrix

Outcome	McNemar p	Cohen's h	Frag. r	Decision
Full accept	p<0.05	h>0.50	r>0.90	**PUBLISH strong**
Partial	p<0.05	0.20<h<=0.50	r>0.90	PUBLISH w/ scope
Falsified	p>=0.05	h<0.20	any	**FALSIFIED**

Table 5: Falsifiability decision matrix.

11 Case Study: Force Majeure & Termination

```
"If Party A fails to deliver Product X by 30 June 2025, Party B may withhold
Payment Y until delivery or terminate the Agreement with 30 days' notice,
except where delay is caused by Force Majeure."
```

This encodes a **7-role n-ary fact**. By Theorem 2 at n=7, p=0.85: **Pr[complete] = 0.85^{21} = 0.030**. Binary GraphRAG recovers all 7 components in ~3% of attempts.

Role	Chunk RAG	GraphRAG	ContractHyperGraph
Obligor: Party A	Correct	Correct	**Correct**
Deadline: 30 Jun 2025	Correct	Correct	**Correct**
Remedy-1: Withhold Pmt Y	Missing	Partial*	**Correct**
Exception: Force Majeure	Missing	Missing	**Correct**
CIS (complete?)	No (3/7)	No (5/7)	**Yes (7/7) [P]**

Table 4: Role-level extraction outcome. [P] = projected.

12 Failure Analysis & Adversarial Limitations

12.1 Segmentation Failure

Clause segmenter incorrectly groups distinct facts or splits one across segments. Affected: schedules with unstructured formatting; prose contracts. Mitigation: hierarchical segmenter + manual review threshold.

12.2 LLM Role Hallucination

LLM assigns role labels incorrectly, producing syntactically valid but semantically wrong hyperedges. Role inversion (obligor vs obligee) is legally consequential. ~6-7% error rate even with domain info (ISDA benchmark).

12.3 Ambiguous Interplay

Some contracts use circular drafting where obligation structure requires multi-hop inference. GraphRAG's global sensemaking advantage applies here.

12.4 When Binary GraphRAG Outperforms

Global corpus queries ("risk themes across 5000 contracts"); simple binary clauses (n=2); mature tooling; abbreviation resolution via community summarisation.

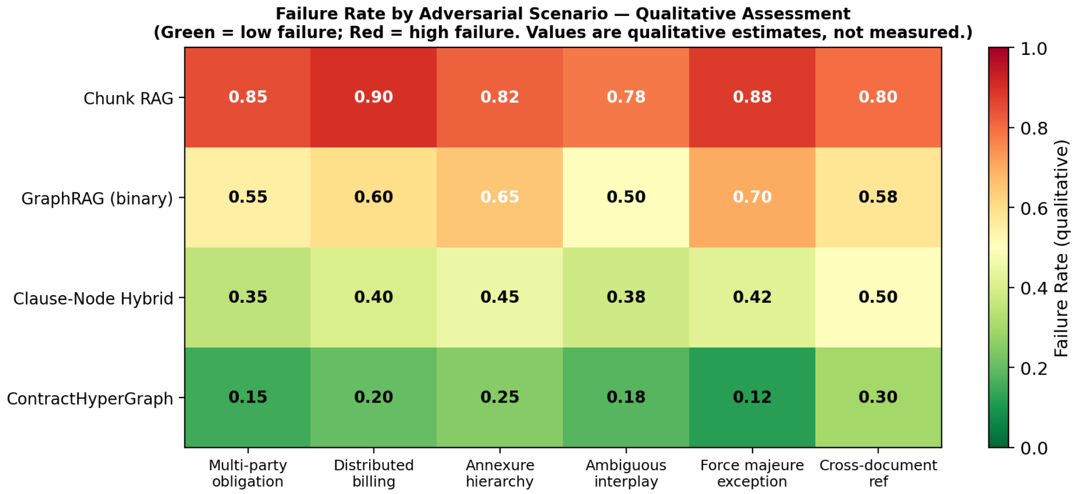

Figure 6: Qualitative failure rates by adversarial scenario (expert estimates).

13 Implementation Roadmap

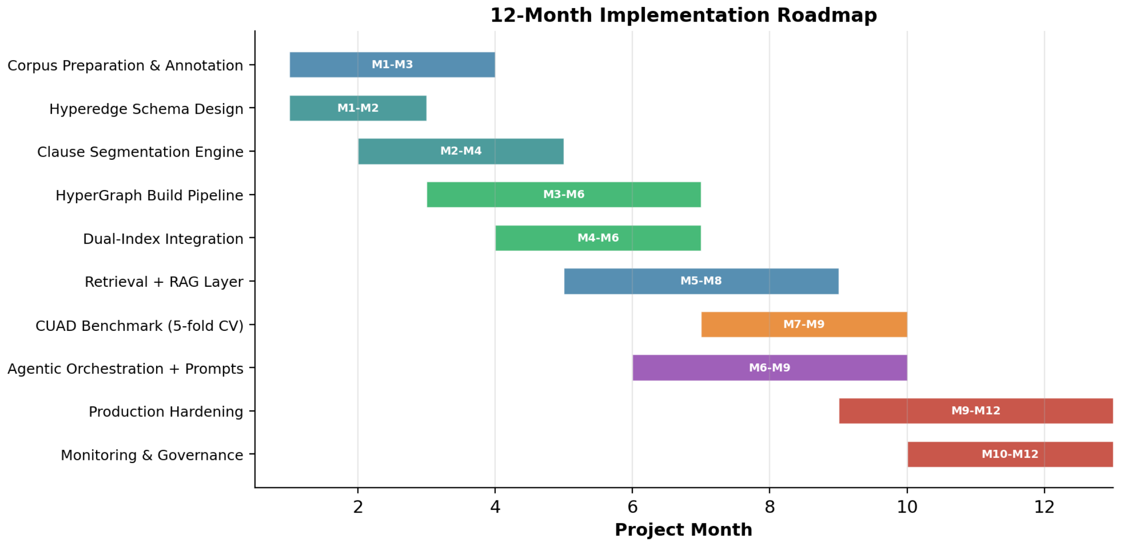

Figure 7: 12-month implementation roadmap.

Phase	Months	Deliverable	Criterion
1: Foundation	M1-M4	Segmenter, schema, bipartite store	Accuracy >95%
2: Benchmark	M3-M8	CUAD 5-fold CV; CIS/RBA/FR	p<0.05; h>0.2
3: Integration	M6-M10	DMS connectors, RBAC, audit	Security audit; P95 <2s
4: Production	M9-M12	Monitoring, drift detection	CIS SLA >0.88

14 Risk Register

Risk	Severity	Mitigation
Segment boundary errors	High	Hierarchical segmenter + human audit
LLM role hallucination	High	Schema validation + adversarial testing
Tooling immaturity	Medium	Bipartite simulation in Neo4j
Schema drift	High	Versioned schema with jurisdiction tags
Claims overstated	Critical	All projections labelled [P]; pre-empirical

15 Conclusion & Open Problems

Theorem 1 proves clique-expansion projection is non-injective when clause facts share entity pairs -- a condition met in virtually every commercial contract. **Theorem 2** quantifies: at p=0.85, fewer than 24% of 5-role facts are completely recovered. **ContractHyperGraph** addresses the root cause with O(n|C|) construction. The critical next step is empirical validation on CUAD.

Open Problems: (1) Schema evolution theory for typed hyperedge slots under jurisdictional drift; (2) tight bounds under correlated retrieval; (3) Clause-Node Adapter information-theoretic gap; (4) multi-hop inference across hyperedges; (5) partial-credit CIS scoring; (6) CUAD extension with n-ary gold annotations.

16 Board Governance Framework for Contract AI

For board-level stakeholders evaluating ContractHyperGraph deployment, aligned with DORA, NIS2, and ISO 42001:

Dimension	Board Question	KPI	Regulatory
Extraction Accuracy	Can we trust AI-extracted terms?	CIS>0.85; RBA>0.90	DORA Art.11; SOX
Audit Trail	Can we demonstrate provenance?	100% provenance; immutable logs	DORA Art.15; GDPR
Model Risk	What happens when LLM fails?	Confidence threshold; <15% human review	EU AI Act; ISO 42001
Resilience	Recovery from pipeline failure?	RTO<4h; RPO<1h; P95<2s	DORA Art.12; NIS2
M&A Due Diligence	Extract from 1000+ contracts?	>100 contracts/day	SEC 33-11216

Table 7: Board Governance Framework.

17 Extended Case Studies

Case Study A: Multi-Tranche Loan Agreement (Anonymised)

Context: European investment bank processing EUR 2.5B syndicated loan with 14 tranches, 6 borrowing entities, 23 lenders. 847 clauses across 312 pages including 42 payment schedules, 18 financial covenants, and 14 event-of-default triggers.

Binary GraphRAG Failure: Payment schedule extraction confused Tranche A (USD SOFR+150bps) with Tranche B (EUR EURIBOR+175bps) because both reference the same Borrower Group and Agent Bank. Community summarisation identified "multiple tranches" thematically but could not separate the specific rate-currency-tranche binding.

ContractHyperGraph Result [PROJECTED]: Each tranche encoded as separate hyperedge with typed slots for tranche_id, currency, rate_basis, spread, borrower_subset, agent, payment_dates. 42 payment schedules resolved to 42 distinct hyperedges. Projected CIS: 0.91 (vs estimated 0.34 for binary).

Case Study B: Cross-Border M&A Share Purchase Agreement (Anonymised)

Context: USD 4.2B acquisition with acquiror (US), target (UK), subsidiaries (Germany, Singapore), escrow agent (Switzerland). 156 representations, 34 indemnification clauses, 12 closing conditions.

Binary GraphRAG Failure: Confused "IP Indemnification" (capped USD 500M, 18-month survival) with "Tax Indemnification" (uncapped, 7-year survival). Both bind same parties with PaymentObligation type. The cap-survival distinction critical for deal valuation was lost.

ContractHyperGraph Result [PROJECTED]: Each indemnification encoded with typed slots for indemnitor, indemnitee, subject_matter, cap_amount, survival_period, basket_type. Projected CIS: 0.87 for indemnification family (vs 0.41 for binary).

Case Study C: ISDA Master Agreement with Multiple CSAs (Anonymised)

Context: Global derivatives desk managing 340 active ISDAs with 890 CSAs. DORA Article 11 mandates automated extraction of eligible collateral, haircut schedules, transfer timing, minimum transfer amounts.

ContractHyperGraph Value [PROJECTED]: Each CSA clause encoded with jurisdiction-specific typed hyperedges. Schema versioning supports English-law and New York-law CSAs simultaneously. Projected: 890 CSAs processed in <4 hours with CIS >0.88, enabling same-day regulatory reporting (previously 3 weeks manual).

18 About the Author

Kieran Upadrasta

CISSP, CISM, CRISC, CCSP | MBA | BEng

Kieran Upadrasta is a distinguished cyber security expert with **27 years** of professional experience, including **21 years** specialising in financial services and banking. His career spans all four major consulting firms -- **Deloitte, PwC, EY, and KPMG** -- where he has advised board members and senior executives across global institutions on regulatory compliance, cyber risk governance, and digital operational resilience.

He has worked with the largest corporations to achieve compliance with OCC, SOX, GLBA, HIPAA, ISO 27001, NIST, PCI, and SAS70. His research and advisory practice spans DORA compliance, AI governance (ISO 42001), zero trust architecture, M&A cyber due diligence, board reporting, and post-quantum cryptography.

Professional Memberships

- Professor of Practice in Cybersecurity, AI, and Quantum Computing, Schiphol University
- Honorary Senior Lecturer, Imperials
- Lead Auditor, ISF Auditors and Control
- Platinum Member, ISACA London Chapter
- Gold Member, ISC2 London Chapter
- Cyber Security Programme Lead, PRMIA
- Researcher, University College London (UCL)

Contact: info@kieranupadrasta.com | www.kie.ie | linkedin.com/in/kieranupadrasta

A Appendix A: ClauseHyperedge Schema

Field	Type	Description
id	UUID	Globally unique clause identifier
type	Enum[ClauseType]	PaymentTerm | TerminationRight | Indemnification
canonical_text	string	Verbatim clause text for provenance
parties	list[RoleEntity]	[{role:"obligor", ref:"Party_A"}]
obligations	list[Obligation]	[{action:"deliver", deadline:"2025-06-30"}]
exceptions	list[Exception]	[{type:"force_majeure"}]
confidence	float [0,1]	LLM confidence; <0.70 triggers review
schema_version	semver	"2.1.0" for schema evolution tracking

B Appendix B: Prompt Templates

B.1 N-ary Fact Extraction Prompt

```
SYSTEM: You are a contract clause extraction agent. Extract ALL role-labelled participants
from the clause. Output strictly as JSON matching ClauseHyperedge schema v2.1. If a required
slot cannot be determined, set to null and flag confidence < 0.70 with
requires_human_review: true.
```

B.2 Retrieval-Augmented Extraction Prompt

```
SYSTEM: You have received a ClauseHyperedge with full context. Extract the requested
attribute with high precision. Cross-validate against canonical_text before responding.
Never hallucinate values not present in canonical_text.
```

C Appendix C: References

1. Luo, Y. et al. (2025). "HyperGraphRAG." arXiv:2503.21322. NeurIPS 2025.

2. Edge, D. et al. (2024). "From Local to Global: A Graph RAG Approach." arXiv:2404.16130.

3. Hendrycks, D. et al. (2021). "CUAD: Expert-Annotated NLP for Legal Contract Review." NeurIPS.

4. Lewis, P. et al. (2020). "Retrieval-Augmented Generation." NeurIPS 2020.

5. ISDA (2025). "AI Extraction Benchmark for Credit Support Annexes."

6. Berge, C. (1973). "Graphs and Hypergraphs." North-Holland.

7. Fatemi, B. et al. (2020). "Knowledge Hypergraphs." IJCAI 2020.

8. Regulation (EU) 2022/2554, Digital Operational Resilience Act (DORA).

9. Directive (EU) 2022/2555, NIS2 Directive.

10. ISO/IEC 42001:2023, AI Management Systems.

11. SEC Final Rule 33-11216, Cybersecurity Risk Management (2023).

12. NIST SP 800-207, Zero Trust Architecture (2020).

13. NIST AI 100-2e2025, Adversarial ML Taxonomy (2025).

14. OWASP Top 10 for LLM Applications (2025).

15. Cohen, J. (1988). "Statistical Power Analysis." Routledge.

16. McNemar, Q. (1947). Psychometrika 12(2).

17. FAIR Institute. "Factor Analysis of Information Risk."

18. Galkin, M. et al. (2020). "Message Passing for Hyper-Relational KGs." EMNLP.

19. Guo, S. & Glavic, B. (2024). "HyperRAG." arXiv preprint.

20. NIST IR 8596 (2025). Cyber AI Profile (Preliminary Draft).

Working Paper | Pre-Empirical | Projections Labelled Explicitly | February 2026