AI Driven Compliance and Secure Data Sharing: Integrating Blockchain and Kafka for Cross Border Media Collaboration

MSc in Software Engineering

Rohan Sikder

Supervisor: Dr. Salim Saay

August 2025

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

Introduction

Problem Context

- Cross-border media collaboration requires **secure and compliant data sharing**.
- Regulations like **GDPR and ePrivacy** impose strict rules on international data transfer.
- Non-compliance risks → **legal penalties, reputational damage, disrupted workflows**.

Research Gap

- Existing solutions study **AI, Blockchain and Kafka in isolation**.
- Lack of **integrated, real-time systems** for automated compliance in data-in-motion.

Challenge & Ambition

- Develop a framework that simultaneously achieves:
  - **Real-time enforcement** of data protection rules.
  - **Transparency & auditability** (Blockchain).
  - **Scalability & low latency** (Kafka).
- Ambitious integration of **AI (legal interpretation)**, **Kafka (event streaming)** and **Blockchain (auditability)** in one prototype.

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

Aims and objectives

Aim

- To design and evaluate a **compliance-aware data sharing framework** for cross-border media collaboration.

Objectives

- Automate **real-time compliance checks** using AI.
- Employ **Kafka** for scalable, low-latency event streaming.
- Use **Hyperledger Fabric** blockchain for **immutable audit logging** and **smart contract enforcement**.
- Demonstrate the framework through a **prototype simulation** of international data exchanges.
- Evaluate the system based on **compliance accuracy, latency, throughput, transparency and security**.

Research Questions

AI & Compliance

- How can AI be used to interpret and automate enforcement of international data privacy regulations in real-time media exchanges?

Blockchain & Smart Contracts

- What are the benefits and limitations of using blockchain smart contracts (Hyperledger Fabric) for automated legal compliance?

Kafka & Data Transmission

- How effective is Apache Kafka in enabling scalable, low-latency, and secure cross-border data transmission?

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

Review of Relevant Literature

Related Work 1 – AI for Compliance

- NLP models (e.g., LegalBERT) help interpret regulations and automate compliance checks.
- Most research is **static (documents, contracts)** rather than **real-time streaming decisions**.

Related Work 2 – Blockchain for Compliance

- Hyperledger Fabric enables **immutability, auditability and smart contracts**.
- Legal tension with GDPR (e.g., “right to erasure”) remains unresolved.

Related Work 3 – Kafka for Real-Time Data

- Kafka widely used in regulated industries for **high-throughput, low-latency streaming**.
- Offers security features but **lacks native compliance enforcement**.

Research Gap

- Current solutions explore **AI, Blockchain, and Kafka in isolation**.
- No integrated, real-time framework that **enforces compliance automatically on data-in-motion**.

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

Methodology

Method Used

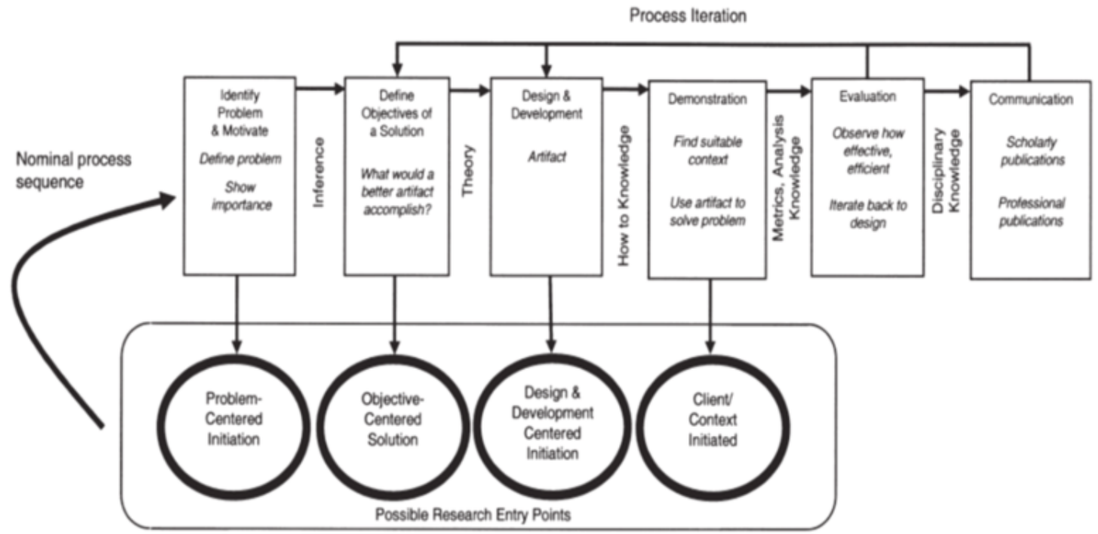
- Adopted a **Design Science Research** approach: problem identification → artefact design → evaluation.

Research Design

- Built a **prototype system** integrating AI, Kafka, and Hyperledger Fabric.
- Simulated cross-border media data transfers (e.g., Ireland -> Germany/Belgium).

Deployment & Evaluation

- Prototype tested in controlled scenarios: compliant transfer, missing consent, jurisdiction block
- Compared performance using metrics: **accuracy, latency, throughput, auditability and security**.



Primary Research & Results

Comparison from Literature Review

- Existing approaches looked at AI, Blockchain, and Kafka **individually**.
- No prior work demonstrated a **combined, real-time compliance framework**.

Requirements Collected

- **Functional**: Ingest events, compliance decisioning, on-chain logging, allow/block.
- **Non-Functional**: Low latency, high throughput, resilience, auditability.
- **Security**: TLS encryption, authenticated identities, role-based access.

Architecture Designed

- **Event-Driven** (Kafka as backbone).
- **AI Compliance Service** for automated policy interpretation.
- **Hyperledger Fabric** for immutable logging & smart contract enforcement.
- **Producer/Consumer Apps** simulate cross-border media partners.

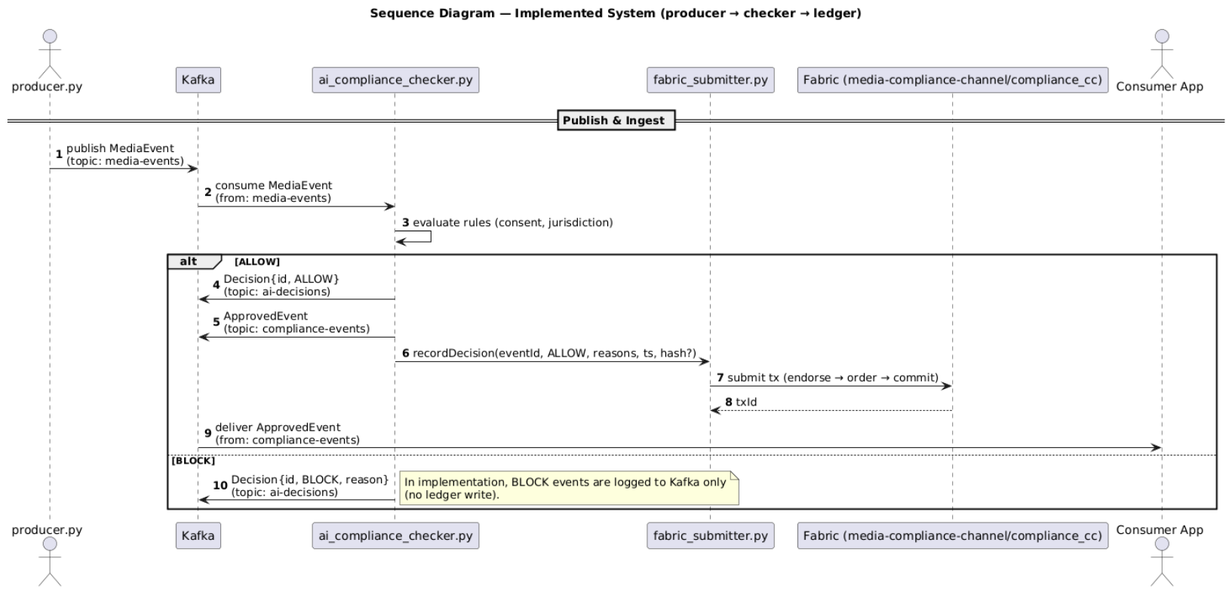
Prototype

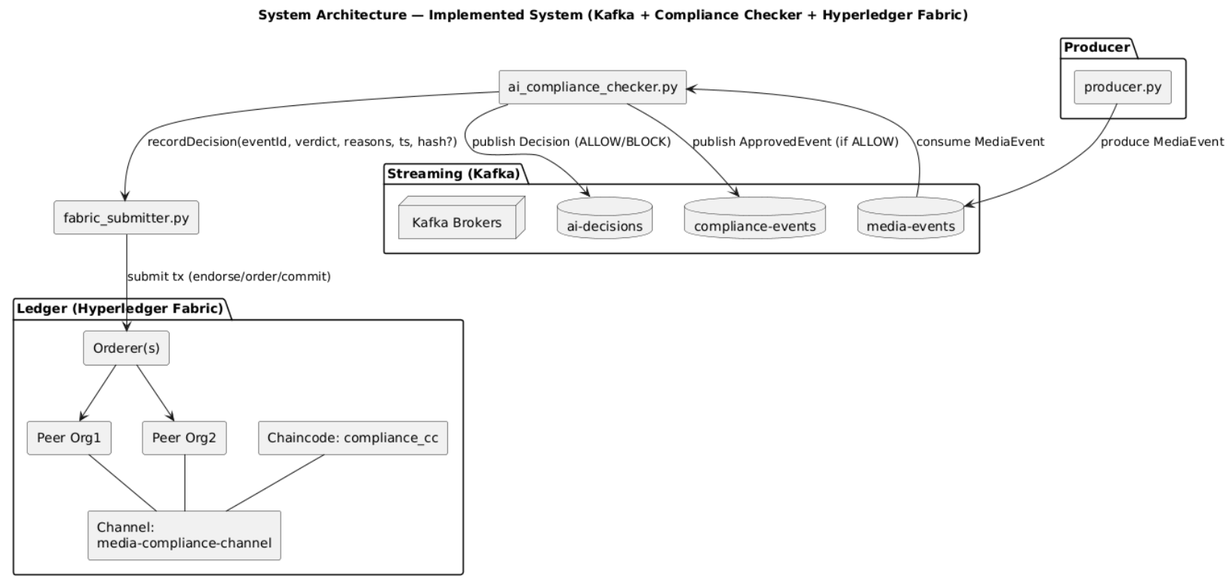
- Implemented a working prototype connecting Kafka → AI → Fabric.
- Tested scenarios: compliant transfers, missing consent, jurisdiction restrictions.

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

Primary Research & Results

- **Prototype Results**
- **Compliance Accuracy**: 100% in simulation (all unlawful transfers blocked, lawful allowed).
- **Latency**:
  - Kafka → AI decision: **tens of ms**
  - Blockchain commit: **~1–2 seconds**
- **Throughput**: Kafka sustained hundreds of events/sec; Fabric write rate slower (few tx/sec).
- **Auditability**: Every decision recorded immutably on-chain, retrievable by event ID.
- **Security & Resilience**:
  - TLS + Fabric MSP identities enforced.
  - No data loss; Kafka replay allowed recovery after failures.





Evaluation

- Prototype evaluated using **simulation of cross-border media data transfers** (Ireland -> Germany/Belgium).
- Test scenarios included:
  - Compliant transfers
  - Missing consent
  - Jurisdictional restrictions
  - Data anonymisation
  - Load and fault tolerance

Evaluation Metrics

- **Compliance Accuracy** – correctness of allow/block decisions
- **Latency** – AI decision latency vs. blockchain commit time
- **Throughput & Scalability** – sustained events per second
- **Auditability** – completeness and retrievability of compliance logs
- **Security & Resilience** – identity enforcement, integrity, and recovery

Findings

- Achieved **100% compliance accuracy** across scenarios.
- **Low-latency decisions** (ms-level) with blockchain overhead (1–2s per commit).
- Kafka demonstrated **high throughput**, Fabric remained the primary bottleneck.
- **Immutable, auditable records** achieved through Hyperledger Fabric.
- System maintained **resilience and integrity** under fault conditions.

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

Practical Session

- **Apache Kafka**: Configured topics (media-events, ai-decisions) to simulate cross-border event streaming.
- **AI Compliance Service**: Rule-based checker to evaluate GDPR-style requirements (consent, jurisdiction).
- **Hyperledger Fabric**: Chaincode deployed to immutably log compliance decisions; query functions for audits.
- **Producers/Consumers**: Python applications simulated media collaborators in different jurisdictions.

Key Demonstrations

- Functional core flow: Producer → Kafka → AI Compliance → Manual Blockchain Logging via CLI.
- Test scenarios executed with accurate allow/block outcomes.
- Fabric explorer and CLI confirmed immutable audit records.

Repository

- Full source code and implementation details available on GitHub:
  👉 github.com/rohansikder/MSc_FYP

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

Conclusion

Achievement of Objectives

- Designed and implemented a **compliance-aware data sharing framework**.
- Integrated **AI, Kafka, and Hyperledger Fabric** into a working prototype.
- Evaluated against compliance accuracy, latency, throughput, auditability, and security.
- Demonstrated feasibility of **real-time compliance enforcement with immutable audit trails**.

Answers to Research Questions

- **AI for Compliance** → AI automated privacy rule enforcement in real time.
- **Blockchain for Compliance** → Fabric provided auditability and trust, though with performance overhead.
- **Kafka for Data Transmission** → Kafka enabled scalable, low-latency, cross-border streaming.

Key Results

- Developed a **proof-of-concept framework** integrating AI, Kafka, and Hyperledger Fabric.
- Demonstrated feasibility of **real-time compliance enforcement** in cross-border media sharing.
- Achieved:
  - **100% compliance accuracy** in simulations.
  - **Low-latency decisions** with moderate blockchain overhead.
  - **Immutable audit trail** ensuring transparency and accountability.
  - **System resilience** under fault conditions.
- Successfully **addressed all research questions** and validated objectives.

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

Future work

- **Advanced AI models** (e.g., LegalBERT, GPT-based) for nuanced policy interpretation.
- **Full Kafka <-> Fabric integration** via Gateway for seamless end-to-end automation.
- **Performance optimisation** of blockchain throughput (batching, consensus tuning).
- **Legal and regulatory validation** through collaboration with compliance experts and regulatory sandboxes.
- **User-facing dashboards** for compliance officers to monitor data flows and audit decisions.

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

References

- Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2008). *A design science research methodology for information systems research.* JMIS, 24(3), 45–77.
- Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). *Legal-BERT: The Muppets straight out of Law School.* Findings of EMNLP.
- Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., et al. (2018). *Hyperledger Fabric: A distributed operating system for permissioned blockchains.* EuroSys.
- Kreps, J., Narkhede, N., & Rao, J. (2011). *Kafka: A distributed messaging system for log processing.* LinkedIn Engineering.
- European Union. (2016). *General Data Protection Regulation (GDPR).* Official Journal of the EU.

(Full bibliography available in dissertation.)

Thank you

Acknowledgements

Supervisor: **Dr. Salim Saay**
MSc in Software Engineering, 2025

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

University of Limerick,
Limerick, V94 T9PX,
Ireland.

Ollscoil Luimnigh,
Luimneach,
V94 T9PX, Éire.

+353 (0) 61 202020

ul.ie