Intelligent Risk Assessment System for Loan Applications

Group 7:

Chen Peng-Wei
Do Quynh Trang
Moo Jia Rong
Pwint Phoo Thaw
Zhang Xiaohan

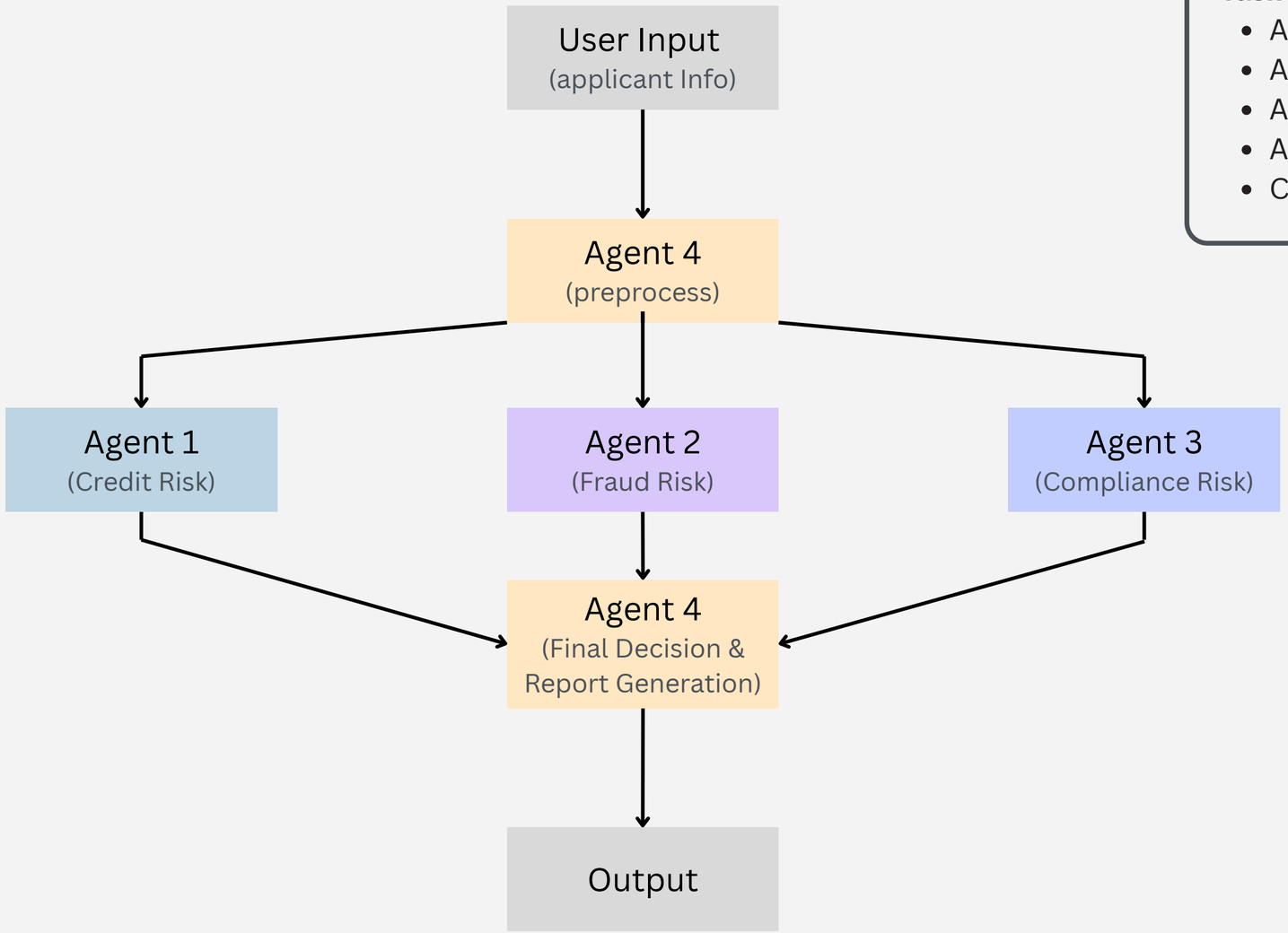
Problem Statement / Use Case

Banks process thousands of loan applications daily. Each application needs checks across credit risk, fraud risk, and regulatory compliance. Lenders must assess creditworthiness, detect fraud, and meet regulations while keeping the process fast. Manual review struggles to balance all these factors consistently across high volumes.

Key Challenges:
Each risk type needs different expertise
Rising delinquencies show current methods miss real risks
Customers want fast answers, banks need thorough checks

Our Solution: A multi-agent AI system where specialized agents handle credit, fraud, and compliance separately, then combine results into one clear decision with full explanation.

Our Proposed Solution



Task Division:

- Agent 1 — Xiaohan
- Agent 2 — Phoo
- Agent 3 — Jia Rong
- Agent 4 & Evaluation— Peng-Wei
- Connecting full pipeline & UI — Trang

Agent 1: Credit Risk Assessment

Role:

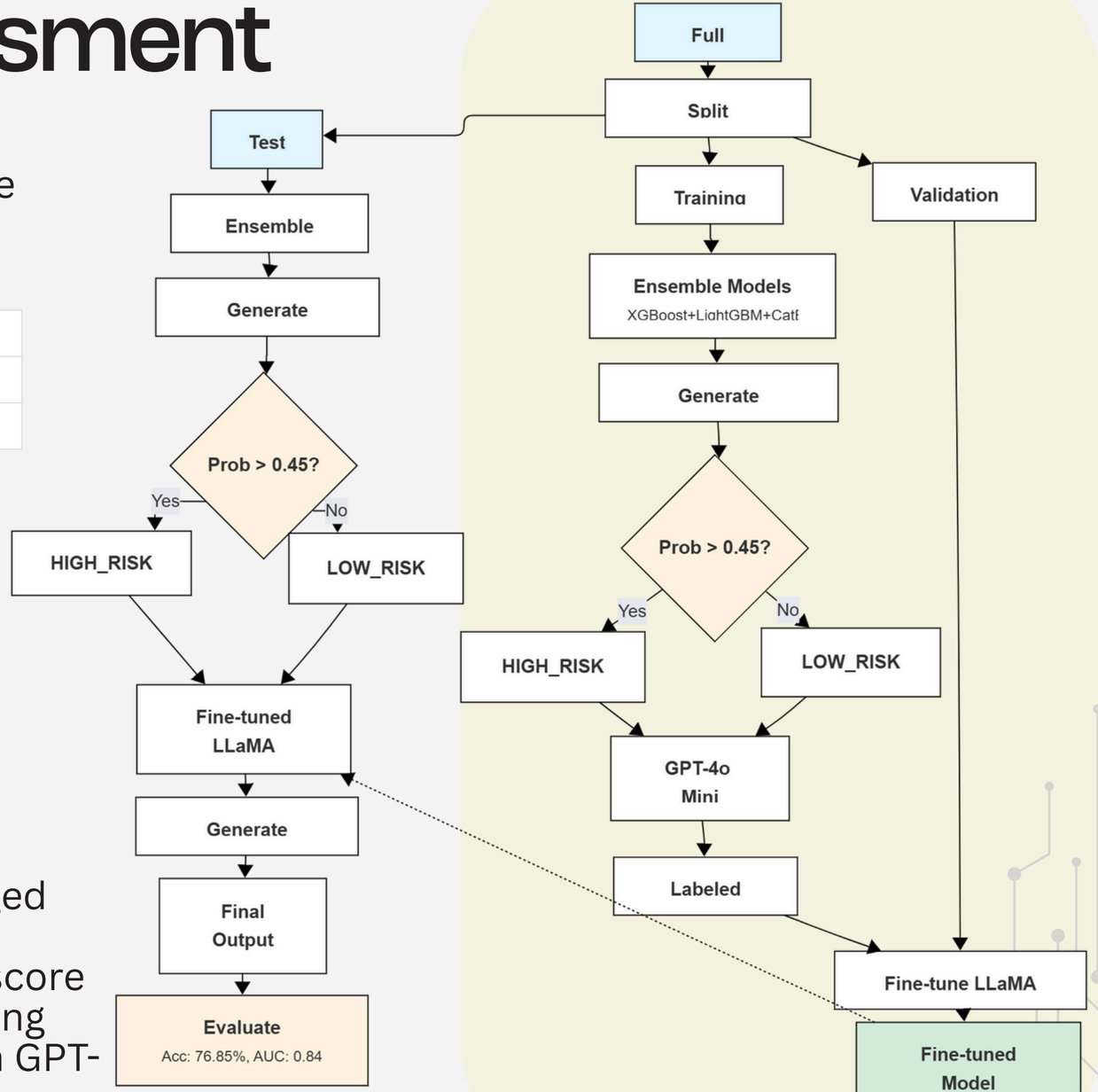
- Predicts borrower credit Risk probability using ensemble ML models, Provides explainable risk assessment with reasoning

Credit Risk Benchmark	Kaggle Link	Credit risk model training
Bank Loan Fraud Detection	Kaggle Link	Fraud detection model training
Regulatory Rules	Internal created	RAG knowledge base construction

Feature	Description
rev_util	Revolving credit utilization rate
debt_ratio	Debt-to-income ratio
monthly_inc	Monthly income
age	Applicant age
open_credit	Number of open credit lines
late_90	Number of 90+ days late payments

Framework:

- Ensemble Models: XGBoost, LightGBM, CatBoost averaged for robust prediction
- Threshold Optimization: Threshold (0.45) maximizes F1 score
- LLM Reasoning: Llama-3.1-8B generates explanations using ensemble predictions and applicant features; trained on GPT-4o-mini labeled reasoning



Agent 1: Credit Risk Assessment

Test Cases: 1,672 cases from Credit Risk Benchmark Dataset
Overall Performance: 76.9% accuracy, 0.84 AUC, 0.78 F1 score
LLM Output Quality: 99.2% coherence, 75 words average length

Reasoning Quality: Rule-based coherence detection using sentiment-risk alignment checks; regex pattern matching for feature coverage analysis; statistical word count distribution.
Explainability Alignment: Compared ensemble feature importance scores (XGBoost gain, LightGBM split count, CatBoost importance) against LLM mention frequencies to identify explanation gaps.
Confidence Calibration: Binned predictions by confidence intervals and compared stated confidence against actual accuracy to quantify systematic under-confidence patterns.

Prediction Accuracy

- TP: 669, TN: 616
- FP: 220, FN: 167
- Conservative bias (1.32:1 FP:FN ratio)

Reasoning Quality

- 4.07 features per explanation
- 99.2% coherence rate
- 13 cases flagged as incoherent
- 10 use valid contrastive structure

Confidence Calibration

- Claims 62% confidence
- Achieves 76.9% accuracy
- Gap: 14.7 points

Explainability Alignment

- Top 3 features: 95% coverage
- Age: 5.0% mention (68.75 importance)
- Credit Lines: 21.5% (62.55 importance)
- Strong alignment for primary drivers (Income, Credit Util, DTI)
- Age and Credit Lines underrepresented

Agent 2: Fraud Risk Assessment

Role:

- Identify false information and fraudulent behavior in loan applications

Check	Description
Application Behavior	Flags rapid or unusual application actions that may signal manipulation or automated submission.
Location of Application	Detects non-local or inconsistent application locations indicating possible identity misuse.
Account Activity	Reviews recent banking behavior for unusual or sudden transaction patterns.
Blacklists	Detects applicants appearing on internal/external fraud lists.
Past Financial Malpractices	Flags known prior fraud/misconduct cases tied to the applicant.
Consistency in Data	Identifies if data is inconsistent across documents and systems.
Previous Loans	Reviews loan history for stacking patterns.

Agent 2: LLaMA-3.1-8B with LoRA

Framework:

- **Statistical Risk Model**: Logistic regression generates calibrated fraud probability using engineered features (e.g., ApplicationBehavior, Blacklists, ConsistencyinData, PreviousLoans)
- **Threshold**: Fraud classification based on optimized cutoff (0.50) for balanced precision-recall and reduced false negatives (missed frauds)
- **LLM Reasoning Layer:** LLaMA-3.1-8B with LoRA produces case-specific fraud explanations based on model signals and feature evidence

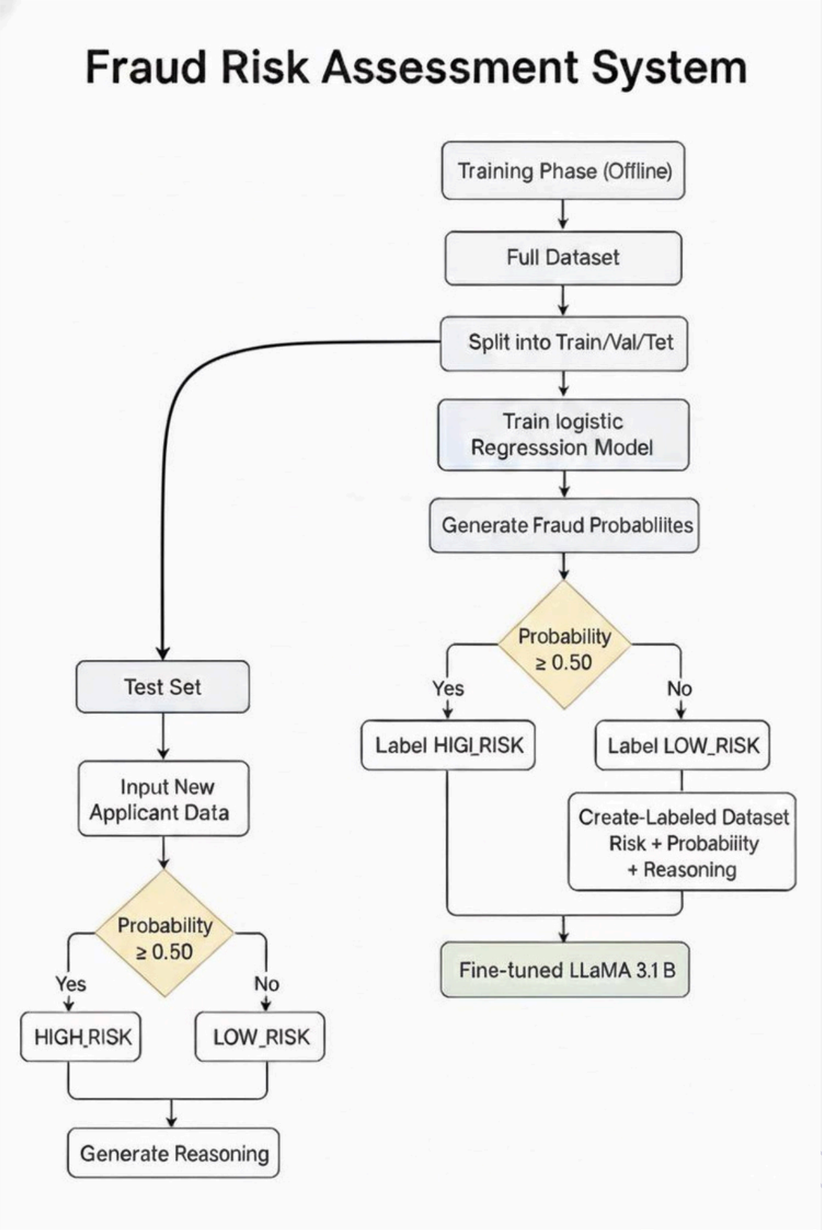
```
================================================================================
INFERENCE COMPLETED
================================================================================

Accuracy: 148/200 = 74.00%
Coverage (LOW/HIGH assigned): 100.00%

Confusion Matrix:
True Positives: 74
True Negatives: 74
False Positives: 30
False Negatives: 22

Precision: 71.15%
Recall:    77.08%
F1 Score:  74.00%

ROC-AUC: 0.8220
```



Agent 2: Evaluation

Evaluation Objective: Ensure Agent 2's fraud reasoning is grounded in provided features, avoids hallucination, and maintains clarity when justifying HIGH/LOW risk.
Test Case: 200 samples
Evaluation Criteria: Alignment with fraud signals, Logical consistency with prediction, Clarity and precision in reasoning
Scoring Scale: 1–5 per dimension (Max total: 15)
Justification Quality: Mean Alignment, Consistency & Clarity scores 10.720 for all checks

LLM Judge: Qwen2.5-14B-Instruct

```
--- Evaluation Complete ---

Average Qwen Judge Scores (Max 15 Total):
Score_1_Alignment       3.340
Score_2_Consistency     3.495
Score_3_Clarity         3.885
Score_Total_Judge      10.720
dtype: float64
```

Agent 3: Regulatory Compliance Assessment

Role:

- Gatekeeper for non-negotiable compliance checks, preventing financial and legal risks

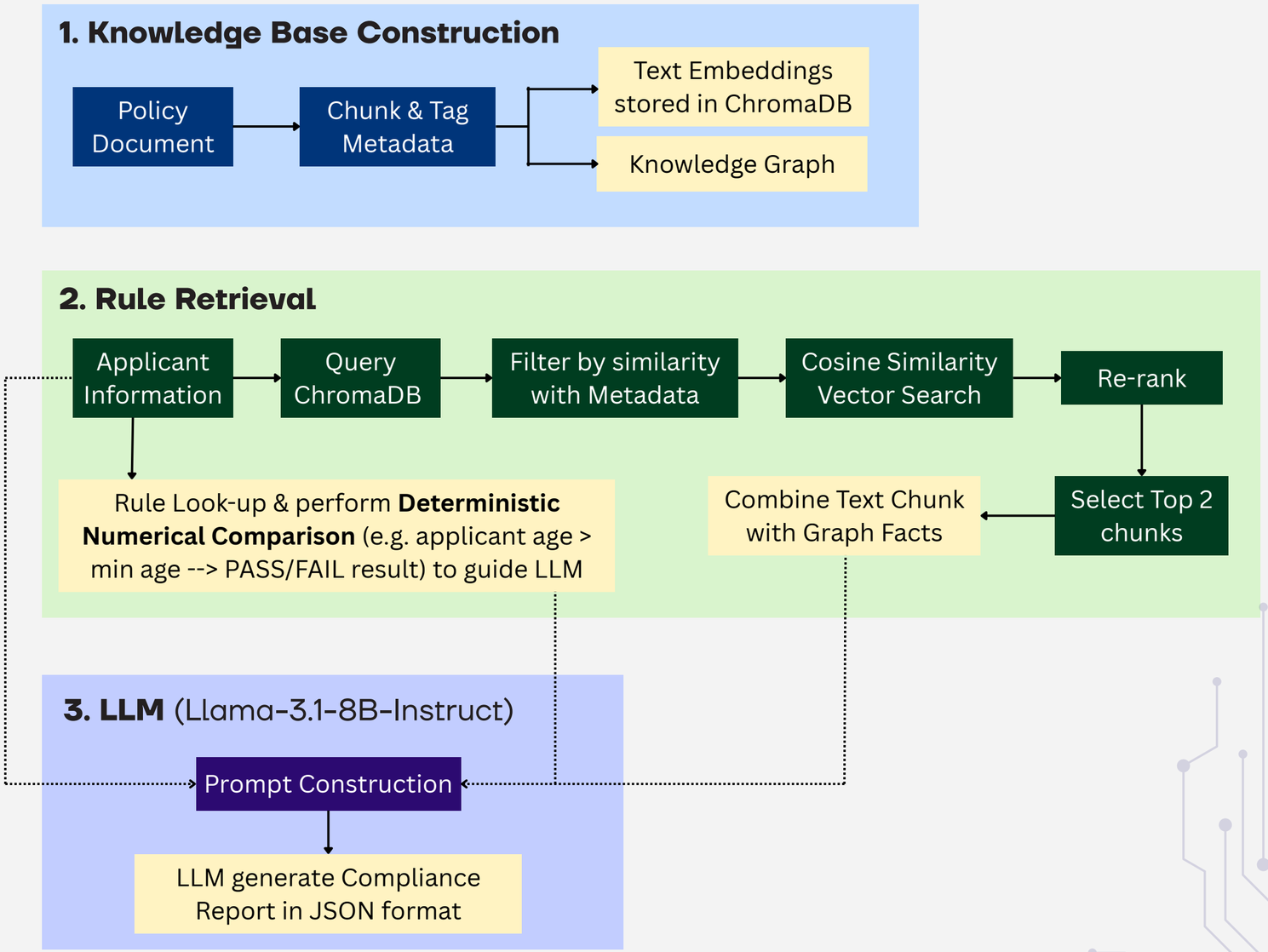
Scope:

Check	Description
Debt-to-income (DTI)	Verifies applicant's DTI ≤ max for loan type
Credit Score	Verifies score ≥ minimum for DTI tier
Income / Employment	Confirms applicant employment
Minimum Age	Confirm that applicant's age meets legal age requirement (18)
Age at Loan Maturity	If applicant age at loan maturity exceeds retirement age (65), additional checks required

Agent 3: Hybrid Graph RAG

Framework:

- **Knowledge Graph:** Structures complex textual rules to aid LLM comprehension.
- **Deterministic Logic:** Evaluates numeric thresholds using Python, ensuring accurate quantitative checks.
- **LLM Reasoning:** Produces faithful, explainable justifications using retrieved textual rules and graph facts; relies on deterministic results for numeric checks and performs reasoning for qualitative checks (Employment).



Agent 3: Evaluation

Test Cases: 173 cases covering all rule combinations
Overall Compliance Decision Reliability: 98.84%
Rule Retrieval: 100% Correct rule in top 2 retrieved chunks
LLM Justification Quality: Mean Accuracy, Relevance & Faithfulness scores above 4/5 for all checks

Compliance Check	Conclusion Accuracy	Recognised correct Textual Rule out of Top 2 provided	Recognised correct Graph Fact to apply	Justification Reasoning Quality (Judged by Qwen2.5-14B-Instruct)		
				Accuracy	Relevance	Faithfulness
DTI	100.00%	**95.38%**	**92.49%**	**4.3**	**4.4**	**4.2**
Credit Score	100.00%	**72.83%**	**82.09%**	**4.23**	**4.25**	**4.19**
Income / Employment	87.28%	100.00%	98.84%	4.63	4.63	4.63
Min Age	100.00%	100.00%	100.00%	5	5	5
Age at Maturity	98.84%	100.00%	98.84%	4.96	4.96	4.96

Deterministic Layer Guarantees Numerical Accuracy

- DTI, Credit Score, Age checks achieved near- or 100% factual accuracy → Deterministic logic eliminates LLM arithmetic errors in numeric thresholds.

Complexity of Nested Policy Rules (Credit Score & DTI)

- LLM occasionally failed to identify the correct rule / graph fact to use for DTI & Credit Score, likely due to the deeply nested, multi-condition logic.
- *E.g. “ If the DTI is less than or equal to 36%, the minimum credit score required is 620. If the DTI is more than 36% and less than or equal to 45%, the minimum credit score required is 680.”*

Agent 4: Decision and Report Generation

Multi-agent LangGraph loan review producing final decision and LLM reasoning.

1. Run Agent 1 to 3 and Collect All Results
Agent 1: Credit analysis + default probability
Agent 2: Fraud analysis + fraud probability
Agent 3: Compliance status + failure details

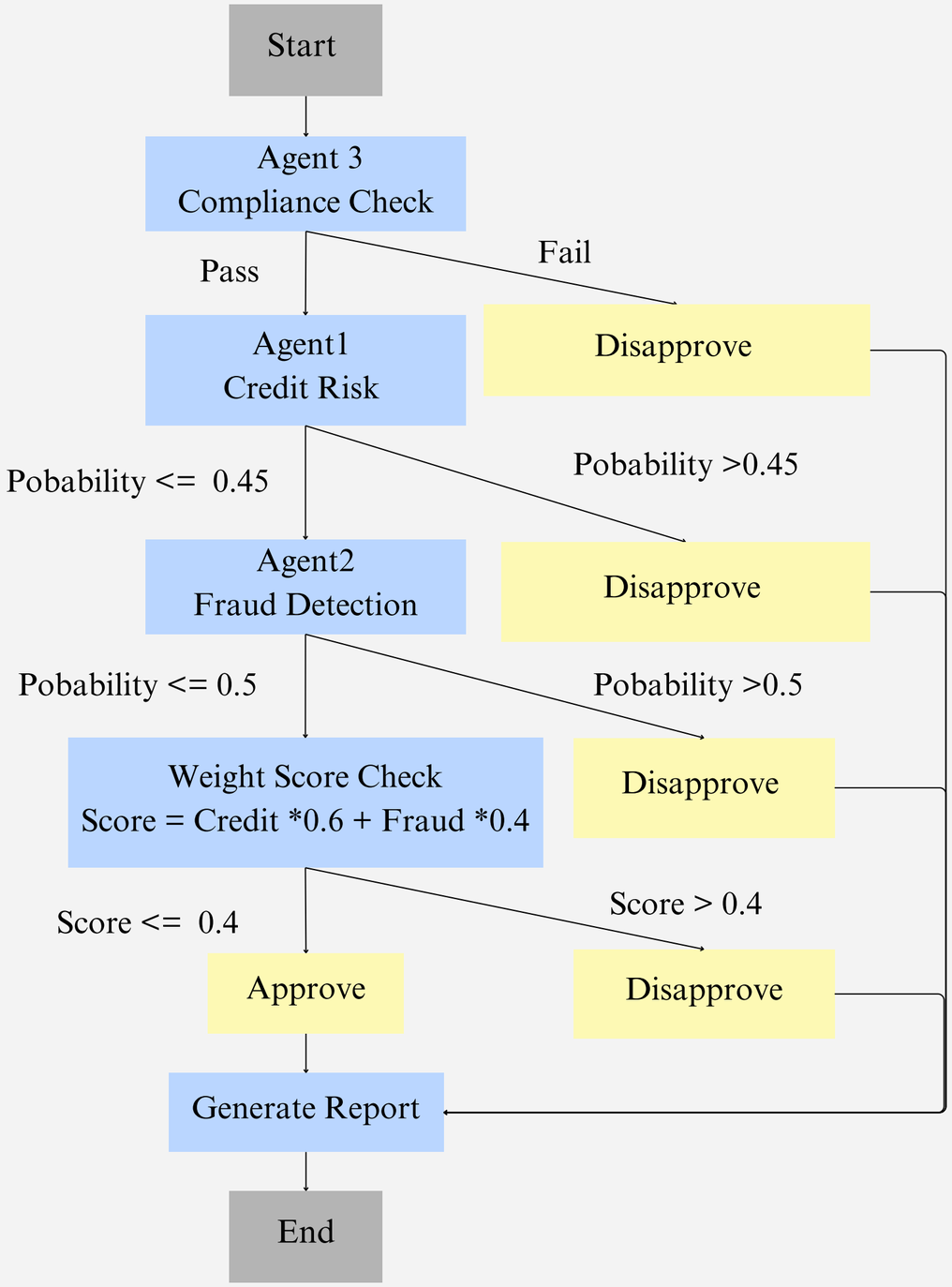
2. Apply Rule-Based Logic
Determine final decision (Approve/Disapprove)
Generate decision reason

3. LLM Synthesis (Agent 4)
Input: All agent results + decision
Process: **Llama-3.1-8B-Instruct** generates reason summary

All agent will run during the reporting phase.

Decision-making phase:
✓ Agent 3 executes → FAIL
✗ Agent 1 skips
✗ Agent 2 skips
→ Decision: Disapprove

Reporting phase:
✓ Rerun Agent 1 (for feedback)
✓ Rerun Agent 2 (for feedback)
→ Generate full report



Agent 4: System Architecture & Output

Group1 : Agent 1 (Independent)
Llama-3.1-8B Base (~7.5 GB)
LoRA Adapters + Ensemble Models

Group2 : Shared Model (Agents 2, 3, 4)
Llama-3.1-8B-Instruct (4-bit, ~6 GB)
Agent 2: + LoRA Adapters (~0.5 GB)
Agent 3: Direct use of base model
Agent 4: Direct use of base model

Total Memory: ~16 GB (vs ~27 GB without sharing) Memory Savings: 41%

Output

Applicant input data + Agent123result +

"agent4_final_report": {
"application_summary": "Application ID: 2, Age: 35, Purpose: home, Credit Score: 800, Debt Ratio: 0.20",
"final_decision": "Approved",
"recommendation": "Approved. The application was strong due to the overall low risk profile presented by the applicant, with Agent 1 (Credit Risk) passing with a low default probability of 0.1922, indicating responsible credit management and a solid financial foundation. Agent 2 (Fraud Risk) also passed with a low fraud probability of 0.0200, showing a typical first-time borrower profile with no recorded malpractices. Agent 3 (Compliance) passed without any failure details, further supporting the applicant's reliability.",
"reasoning_trace": [
"1. Compliance: PASS",
"2. Credit Risk: PASS (prob: 0.1922, threshold: 0.45)",
"3. Fraud Risk: PASS (prob: 0.0200 threshold: 0.5)",
"4. Weighted Score: PASS (0.1233 threshold: 0.4)",
"5. Final Decision: Approved"
]
}

LLM generated

Agent 4: Evaluation

Test Cases: 32 data covering all result combination
(Agent1 Pass/ Fail + Agent2 Pass/ Fail + Agent3 Pass/ Fail + Weighted score Pass/Fail, total 8 combintaions)
Decision Consistency: 100 %
(reasoning_trace vs LLM recommendation, final_decision vs LLM recommendation)
Recommendation Faithfulness: 100%
All approved Agent is not misrepresented and mention all failed agents in LLM recommendation.
FAIL Case Completeness :100%
First sentence of the recommendation mentions all the actual failed agents.
3 Agent Mention Frequency in Recommendation: 100%

LLM Judge: Qwen2.5-14B-Instruct

Agent	Mention_Rate	Avg_Score	Avg_Unfaithful_Score	Unfaithful_Cases
Agent 1	100.0%	5	N/A	0
Agent 2	100.0%	5	N/A	0
Agent 3	100.0%	4.81	3.00	3

Extract Agent3 Failed case : Only mention one of the failed reason
A1 Score: 5, A2 Score: 5, A3 Score: 3 Justification: Agent 3's summary incorrectly states that the application was disapproved due to a debt-to-income ratio exceeding the maximum allowed DTI of 45%. While this is one of the reasons for failure, the primary reason given in the original reasoning is that employment verification is required for income verification, which is not mentioned in the recommendation.

Future Work

1. Adaptive Hierarchical Evaluation

- Current: Agent 3 (compliance) runs first → most costly.
- Future: Use Agents 1 & 2 for pre-screening instead to improve efficiency. This further enable introduction of tiered compliance system, where Agent 1 & 2 outputs determine whether a case undergoes “standard” or “strict” compliance review.

2. Accuracy Enhancement (Agents 1 & 2)

- Targeted fine-tuning on edge cases.
- Broaden training data for greater robustness and generalization.

3. Explainability-Aware Reasoning (Agents 1 & 2)

- Integrate SHAP feature importance to guide LLM attention toward high-impact features during reasoning generation.

4. Expanded Compliance Coverage (Agent 3)

- Incorporate additional compliance checks and automate policy ingestion and rule maintenance for scalability.

5. Structured Rule Representation (Agent 3)

- Convert nested policy logic into structured forms (e.g., JSON / logical trees).
- Enable LLM to reason over explicit IF–THEN conditions, reducing ambiguity in complex, tiered rules (current graph facts capture only relationships e.g. Home_Loan → DTI_36_to_45 → Min_CreditScore_680)

THANK YOU