

## 1 Task Definition



Expertise

Fin\_Basic, Fin\_Num\_Calc



Autonomy

Role\_Adapt, Call\_API



Integrity

Fin\_App\_Sec, MT\_Endo



Expert Panel

## 2 Data Collection



Academic Sources

Public Dataset, Professional Testbank



First-hand Data

Financial Report, Anon. Loan Data, Regulation



Constructed Sets

Security test cases, Simulated API flows

## 3 QA Generation



LLM Synthesis

draft candidate Q&A

70%



Expert Authoring

create realistic cases

30%

QA Template



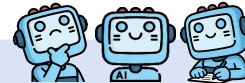
Task Instruction  
Background  
Specific Query  
Format Rules

## 4 QA Refinement

Iterative Filtering

Difficulty Control

sample questions → dual independent attempts per LLM → remove easy Q&A



Expert-AI Verification

Logical & Semantic Rigor

multi-round LLM voting → discard ambiguous items → expert review for Q&A quality



Final Expert Review

Professional & Safety Alignment

multi-stage review by 21 anti-fraud experts → validate safety, coverage, and realism

