Case Illustration

WhitePine Financial Holdings: AI as a Cross-Cutting Risk

Fictional illustration prepared for discussion purposes

WhitePine Financial Holdings is a Canadian financial institution with operations spanning retail banking, commercial lending, wealth management and insurance. Over the past two years, WhitePine has embedded AI systems across its core business functions. Three events in 2026 have raised questions the organization has struggled to answer cleanly.

The Credit Adjudication Platform | Machine Learning

In 2025, WhitePine deployed a machine learning model to accelerate small business and consumer lending decisions. The model integrated traditional credit variables with behavioural transaction data, real-time fraud screening and dynamic pricing logic. Unlike a rules-based system, the model incorporated ongoing retraining and periodic parameter updates informed by new data, and its credit assessments evolved as those updates took effect. Approval volumes increased, processing times fell, and management considered the deployment a success.

During a recent period of regional economic volatility, the model began declining applications that WhitePine's own credit policies would otherwise have supported, and pricing new credit facilities at spreads wider than WhitePine's standard risk-based pricing would have produced. Applicants quoted higher rates either went elsewhere or accepted terms more expensive than warranted. Those who were declined received a notice of the decision, but branch staff could not explain the basis for it if asked. Certain industries and communities appeared to be disproportionately affected. Complaints began circulating on social media, and local news media began questioning whether automated decision-making was disadvantaging particular borrowers.

Responsibility for monitoring model performance metrics sat with Model Risk; metrics remained within approved tolerances so no escalation was triggered. Legal and Compliance assessed whether complaint handling continued to meet FCAC standards during the surge. The business interpreted the model's response as appropriate given elevated economic risk. Customer Relations logged the surge in complaints as a service quality issue and routed them to the relevant business lines for response.

Each function observed indicators moving in concerning directions, but no single metric crossed the specific trigger that would have required escalation. The framework had been designed to respond to discrete breaches within risk categories, not to aggregate signals accumulating across them simultaneously.

By the time the issue reached executive management, OSFI and FCAC had already sent information requests.

The Call Centre Assistant | Generative AI

WhitePine's retail Call Centre had introduced a generative AI assistant that consulted the bank's actual policy documents in real time to help agents respond to customer inquiries. But in edge cases such as complex mortgage modifications and insurance claim eligibility, it produced responses that were confident, plausible and subtly incorrect. Call Centre agents, many of whom lacked deep product expertise, had little basis to question outputs that appeared authoritative. Several customers received inaccurate information about their loan terms and entitlements before the pattern was identified. Because interactions were summarized rather than fully retained, WhitePine struggled to determine which customers had received inaccurate information or whether affected customers could be identified systematically.

The Cash Management Tool | Agentic AI

WhitePine's Treasury division had deployed an agentic AI system to manage the bank's intraday cash positions, deciding how much money to hold in different parts of the business and moving funds between business units as market conditions changed, within parameters that had been pre-approved.

The system was authorized to execute liquidity reallocations automatically unless interrupted by authorized personnel. During a sudden market disruption, the system began acting faster than Treasury staff could follow. As conditions deteriorated, several staff concluded independently that the pre-approved parameters were no longer appropriate for the situation unfolding. The Treasury analyst monitoring the system attempted to pause specific transactions. Their manager, unaware the analyst was already acting, issued a different instruction. The head of a business unit whose funds were being reallocated tried to block those movements. The risk function, receiving escalation calls from multiple directions simultaneously, issued its own directive. Simultaneous attempts created conflicting instructions rather than effective control.

Escalation to senior leadership became congested. The organization's escalation framework had been designed for decisions made at human pace, and the system continued to operate throughout.

One More Thing

WhitePine's insurance broker also advised that carriers are introducing AI exclusions across D&O, professional liability, and general liability policies at renewal. Some categories of AI-related exposure WhitePine may have assumed were transferable are no longer insurable at economically viable levels. The board has begun asking whether the organization is accumulating forms of operational and governance risk that markets themselves are becoming unwilling to absorb.

Full Group Discussion: Your Assignment

The board Risk Committee has received separate briefings on all three events. Each business line has described its piece of the problem. The board is not satisfied. No one has connected these events into a coherent account of what WhitePine is actually facing.

As Special Advisor to the Chief Risk Officer, you have been asked to do exactly that. Your CRO has raised the following questions:

1. Do these events represent weaknesses in individual risk functions, or do they expose something more fundamental about how WhitePine is architected to govern risk in an AI-enabled business?

Credit Adjudication Platform

2. What risk categories were active in that scenario?
3. Where was the governance weakness? Was it Model Risk? Compliance? Business leadership?
4. Model Risk confirmed the model's performance metrics remained within approved tolerances throughout. Yet the model was producing outcomes that WhitePine's own credit policies would not support. How is it possible for a model to be performing correctly on every metric that Model Risk was tracking and still produce outcomes the bank did not intend? What does that tell us about what those metrics were measuring?

Call Centre Assistant

5. The credit platform produced unexpected behavior under conditions that might have been anticipated and tested for. The call centre system was operating exactly as designed and still produced wrong outputs. What does that difference mean for how you would approach governing each?

Cash Management Tool

6. Multiple people attempted to override the system simultaneously. None of them succeeded, and their attempts created conflicting instructions. What would have needed to be in place for any one of those override attempts to have been effective?

* * *

7. Which assumptions embedded in WhitePine's existing ERM architecture appear no longer reliable under conditions of large-scale AI deployment?

Breakout Session Discussion: Questions

You remain in the role of Special Advisor to the CRO of WhitePine. Each group has been assigned one of the six questions below. Be prepared to share your key observations with the full group.

Q1 - Credit adjudication: Monitoring and detection

Model Risk confirmed that performance metrics remained within approved tolerances throughout the credit platform event. Yet the model was declining applications and overpricing new facilities beyond what WhitePine's own credit policies would support. What monitoring of model outcomes rather than model performance metrics would have detected this earlier? At WhitePine, who might be best placed to take responsibility for that monitoring, and what expertise would they have needed to assess what they were seeing in real time? What outcomes, behaviours or customer impacts should have been treated as risk indicators even though traditional model metrics remained stable?

Q2 - Credit adjudication: Remediation

WhitePine now has a backlog of potentially affected lending decisions: wrongful declines and overpriced new originations. What would a credible remediation program require in terms of reviewer capacity, access to application files and model decision records, and credit judgment exercised independently of the model under review? What does the FCAC 56-day complaint resolution requirement mean for how quickly that capacity must be mobilized, and what would WhitePine, in hindsight, have been well advised to build into the model deployment to facilitate complaint resolution?

Q3 - Call centre: Detection and remediation

The generative AI assistant was operating as designed and yet customers received subtly incorrect information about their loan terms and entitlements. No threshold was breached and no metric flagged a problem. Unlike the credit adjudication scenario, there is no defined backlog of identifiable decisions to review. What

does this reveal about the limits of traditional monitoring for this type of AI system, and what would identifying and remediating affected customers require? What design-stage decisions would have made it possible to detect this problem earlier and identify which customers were affected? What level of logging, retention and interaction traceability would have been necessary to investigate the issue effectively after the fact?

Q4 - Cash management: Governance of agentic systems

The agentic cash management system operated within its pre-approved parameters throughout the market disruption. Yet Treasury staff could not determine what it was doing or why, override attempts created conflicting instructions, and escalation to senior leadership became congested. What does this reveal about the limits of parameter-setting as a governance mechanism for agentic AI systems? What authorization architecture and human oversight design would WhitePine need before deploying this kind of system at operational scale?

Q5 - Cross-cutting: Accountability

Across all three events, each function operated within its defined lane and WhitePine was still blindsided. Consider both the first line business functions that deployed and operated these systems and the second line oversight functions responsible for the risk framework. Where should accountability for end-to-end AI system outcomes sit in the first line, and what coordination and integration role should the second line play to ensure that what each function sees is assembled into a coherent picture?

Q6 - Cross-cutting: Governance designed to cope with problems in production

In all three events, problems reached customers and regulators before WhitePine's internal governance detected them as enterprise risk events. WhitePine's governance architecture appears to have been designed around the assumption that controls would prevent problems from reaching production. How would WhitePine need to change its governance architecture to cope with AI-related problems in production?

Grand River Valley Bank

You work at Grand River Valley Bank (headquartered in Waterloo ON). GRVB has recently onboarded a new team that focusses on financing technology companies. The target market includes software/SaaS and AI firms. These firms have been through a number of financing rounds but would like some debt to avoid further dilution.

The team has been clear that most of the loans would not be collateralized and would not be generating sufficient cash flow to confidently meet timely interest and principal repayment. The equity investors in these firms are founders with a track record and strong VC partners.

The team will be called the "Tech Banking Team" or TBT and they will bring with them origination bankers and loan specialists, as well as middle and back office support groups.

TBT's value proposition is that, while the loans on their own would be unprofitable, the firm would put its operating deposits with your bank and your wealth group would be invited to speak to their executive team about on-boarding as clients. Their previous bank also had deep relationships with the VC firms and their executives which they hope to bring to GRVB.

GRVB's CEO was personally involved in the team's hire and has publicly and privately stated her excitement and optimism about what the team will bring to the bank. During the last quarterly call with analysts, the CEO and CRO both discussed the new team and stated that GRVB was not changing its risk appetite with this new initiative and that the team's strong risk culture would be an excellent fit with that of the bank.

The second-line risk group (IRM) has formed a team to work with TBT and other 1LOD leaders to start to work through the issues that may arise as they start to do business.

Some background information:

- GRVB's Board has just approved its strategic plan and associated risk appetite.
- The new team would be part of GRVB's Commercial Bank division. The Commercial Bank has run a traditional book of business and has a Moodys equivalent WAR of Ba1.
- Current Board-approved risk appetite metrics were calibrated off of the existing book of business and did not anticipate the acquisition of the new team. The metrics were calibrated to be well inside the risk capacity of the GRVB.

TBT and the head of the Commercial Bank have drafted a risk appetite for the new business. Key highlights include:

- Authorized TBT loans will be no greater than 15% of the authorized commercial portfolio.
- TBT's relationship ROE will meet or exceed the Commercial Bank target of 13%.
- TBT's Chief Credit Officer will be able to approve any loans smaller than $100MM up to a borrower limit of $250MM. Above this amount, the head of TBT and the head of the Commercial Bank will need to jointly approve. Note that credit approvals at GRVB are done within the 2LOD.
- TBT will be exempt from GRVB's credit policies and procedures and subject to their own until such time as they can agree with GRVB on new frameworks.

Your manager in the 2LOD has asked your team to draft a response to this in advance of the meeting. While your manager will be there, this will be yours to present.

The head of TBT and his direct reports will be there, as well as the head of the Commercial Bank. The head of TBT is known as charismatic and charming but highly driven and passionate about his team and their clients. TBT's track record is impressive having delivered ROEs above 25% for the past 3 years and revenue CAGR of 15% over the same period.

Risk Reporting Example:
Undrawn Commitments

Undrawn (or unfunded) commitments are the difference between the authorized loan amount and the amount drawn, where the lender has a legal obligation to fund the undrawn portion at the borrower's request, as long as certain conditions are met.

Common banking and wealth products that have this feature include:

- Revolving credit lines to consumers – these can be secured or unsecured
- Securities-based margin facilities in wealth management
- Revolving credit facilities for commercial borrowers
- Delayed draw term loans
- Letters of credit or letters of guarantee

Regulatory reporting and capital treatment usually broadly tracks the above.

You've been asked to create a report on undrawn commitments. The primary audience for the report will be risk teams in 1LOD and 2LOD, as well as Corporate Treasury, who want to understand the funding and liquidity implications of unfunded commitments.

For now, a weekly report is adequate but management would like the option to produce it on demand or more frequently in times of stress.

Assume that your firm has 3 main business segments: Retail, Commercial and Wealth. Retail is a Canada-only business but Commercial and Wealth have U.S. and Canadian operations. Each segment has its own loan system but your firm has a common set of reference data and a common credit risk system.

- Loan systems data for each segment include authorized and drawn amount by facility for each borrower and by currency. They also have the SIC code for each facility.
- The common credit risk system has credit facilities with ratings and assigned loss-given-default codes. SIC codes are also included in credit data and can be different from the loan data since credit may over-ride the business input.
- Reference data have single name hierarchy, sector classifications, ratings mapping and LGD codes. They also house legal entity information that lists which of your firm's operating entities holds the loan.
- Your Risk Technology team has a legacy risk and aggregation engine that is fed by these data sources and can be modified to calculate the undrawn exposure and aggregate the data according to your specifications.

- A reporting layer will draw from the engine and produce the report.

A simple illustration of the architecture is below.

Propose a risk reporting structure given the information above.

1. What are some of the key questions we need to address with respect to:
    - Scope of the report
    - Granularity vis-à-vis intended audience
    - Cadence of reporting
2. What are some of the BCBS 239 issues that might arise in the development of the report, especially in terms of:
    - Governance and infrastructure
    - Risk data aggregation
    - Risk reporting practices

Undrawn Commitments Risk Reporting Architecture