STRATEGIC AI BRIEF

Northbridge Financial Group

AI Transformation Initiative

Prepared by Luke Kilpatrick

University of Waterloo AI-CTO Certificate Program

July 2026

1. Where Northbridge Actually Stands

Northbridge has three AI systems in production. Only one was a deliberate strategic investment, and none is Custom AI: we have never trained, fine-tuned, or owned a model.

The fraud detection model is genuine Embedded AI: built in-house four years ago, maintained by a six-person ML team, processing 2.1 billion transactions annually. It works. This team should help interview and train the people we hire, not be pulled onto new projects.

The other two systems are Configured AI. Copilot was deployed to 2,200 employees eight months ago at 31% utilization. Low adoption has one hidden benefit: we have not yet hit the issues a missing governance policy surfaces at scale. That is a risk waiting to happen.

Then there is Maple. A customer-facing chatbot handling 4,200 sessions per day on a third-party LLM API, recently subject to a security incident requiring regulatory notification. Maple is what happens when Configured AI ships without governed infrastructure underneath it, and it is a leading indicator of what happens at scale if we go agentic without closing that gap.

The CDO's data quality memo puts a hard constraint on that timeline: three of five internal data systems have material quality issues, with six to eight months to remediate. A model generating confidently wrong output from bad data is worse than no model at all: it carries the authority of automation without the accuracy to justify it.

None of the three proposals is fully executable today. The ML team gets us off home plate; scoring a run means building the team around them, not borrowing from them. Every path starts with hiring, governance, and data remediation before deployment.

2. The Recommendation

My recommendation is Proposal C. Not because it is cheapest or fastest; it is neither. But it is the only path that lets Northbridge change direction as the AI landscape changes underneath it, and it changes every four weeks or less.

A and B look cheaper and faster on a spreadsheet. But both lock us into someone else's architecture just as the environment shifts against it. Signal 1: Microsoft is reviewing enterprise pricing upward 30 to 50 percent, and our EA expires in fourteen months. Signal 2: the EBA is consulting on AI sovereignty, directly affecting our German operations and indirectly the shared infrastructure supporting the UK. Proposal A is exposed to both, Proposal B indirectly. Proposal C materially reduces both exposures, because we control deployment location, model selection, infrastructure sourcing, data pipelines, and compliance logic. It costs more time and money, but fast or cheap usually means the job is not done well.

The first six months are foundation: governance, data remediation, hiring. On hiring, the instinct is to recruit people who have done this before. I would push back. Everyone in AI is building the airplane in flight; today's tooling and regulations will not look the same in eighteen months. Hire two to three senior engineers who can architect and mentor, then build around them with talent selected for curiosity and adaptability over resume keywords.

The acquisition (Signal 3) is worth pursuing but not load-bearing. It buys acceleration: compliance IP and potentially six to eight months off the timeline. If due diligence does not hold, Proposal C still works; it just takes longer. It is an accelerant, not a prerequisite, and only Proposal C can use it: A and B do not own their platforms.

The plan fits exactly in the $15M envelope: $3M foundation phase, up to $10M for acquisition and platform execution, $2M in gated reserve. The ask today is bounded: approve Proposal C, release the $3M foundation capital now, authorize recruitment of at least two senior AI engineers, permit acquisition due diligence, and hold the remaining $12M behind the gates in Section 6. The first production capability arrives in eight to ten months, not three. But when it arrives, we own it.

3. What This Actually Costs

The finance team will compare the three proposals on Year 1 price. That comparison is misleading. Look at what just happened to GitHub Copilot.

On June 1, 2026, Microsoft moved GitHub Copilot from flat-rate subscriptions to token-based billing. Users reported projected bills jumping from $29 per month to $750, and from $50 to $3,000. Those are user-reported projections, not published price increases, but the pattern is the point: economics hidden behind a subsidized flat rate became visible when billing shifted to consumption. That is the direction of travel for platforms on someone else's infrastructure.

That is Signal 1 applied to Northbridge. Proposal A states $4.2M for Year 1 on an EA expiring in fourteen months, with a 30 to 50 percent pricing review announced. Agentic costs behave differently from license costs: each workflow can generate multiple inference calls, retrievals, tool executions, validation steps, and retries, so spend can become highly variable and rise faster than user adoption. Realistic Year 2 cost on Proposal A is $5.5 to $6.3M, and may be conservative.

The real cost of an AI initiative is not the build but the run, and the difference between the proposals is who controls that price. Proposal A: lowest upfront, highest volatility, no leverage at renewal. Proposal B: moderate upfront, but the vendor owns the weights and training pipeline. Proposal C: highest upfront, most predictable run cost over 24 months, because we own the compute, models, and data pipeline.

The metrics to track quarterly: cost per completed workflow, usage and volume trajectory, AI spend against measurable revenue or operational impact, forecast variance, and vendor concentration and renewal exposure. If those drift by Month 10, we adjust before the reserve is committed.

4. The Security Questions We Should Actually Be Asking

The Maple incident was a chatbot security failure, and the security and governance team is right to demand it never recurs. But Maple generated text; these proposals deploy systems that act on live credit, compliance, and underwriting infrastructure. A chatbot hallucinating a wrong answer is a customer service problem. An agent executing a compliance decision on flawed retrieval is a regulatory and financial exposure problem. The risk shifts from performance to control and governance: who is accountable when the system acts on its own reasoning.

We should anchor the program in the NIST AI Risk Management Framework, integrated with our existing cybersecurity controls. NIST provides the structure: govern, map, measure, manage. The OWASP Top 10 for Agentic Applications and MITRE ATLAS provide the threat models for concrete red-team scenarios. The framework governs; the threat models test.

Hallucination is a characteristic of these models, not a bug that gets patched, so in workflows touching money, credit, or compliance, a single agent alone carries too much risk. One supporting control is multi-agent consensus: three agents evaluating independently, on different models, not copies. Agreement advances a result to the next validation stage; disagreement forces recalculation or human review. Consensus is a validation signal, not an authorization control: agreeing models do not prove correctness; errors may be correlated. Deterministic policy controls, audit trails, and human authorization remain the controls of record for credit, compliance, and money. This is banking's dual-control applied to AI: no model's confidence score, and no model vote, moves money or approves credit.

Before production, we define three things:

What we approve: which systems an agent can access, what it can do autonomously, and where deterministic controls and human authorization sit. Customer service suits agent-assisted resolution; credit decisioning needs rules-based architecture with AI-assisted analysis.

What we monitor: incident metrics, hallucination rates by workflow, alignment drift, and telemetry against the NIST and MITRE ATLAS threat models. Red-teaming is a recurring operational investment.

What we own: risk appetite. How much autonomy to grant an AI system acting on financial data is an executive question we bring to the board with a clear recommendation.

We cannot outsource this: A inherits Microsoft's security posture, B the vendor's, C means we own the architecture and the accountability. After Maple, which position do we want to defend to a regulator?

5. Governance Is the Foundation, Not the Finish Line

Northbridge is a bank. Our risk tolerance is lower than a startup's, and moving at startup speed is a sure way to crash into a wall. We want to be Bugs Bunny: smart, tactful, with a plan. Not Wile E. Coyote, ordering from a vendor catalog and praying it does not launch us off a cliff.

Our governance posture today is a draft policy and good intentions: no risk tiering, no standing body, no audit trail for any AI-driven decision. The six-to-eight-month remediation window is not dead time: stand up a governance working group immediately, meeting and documenting biweekly, so the framework is operational and tested before the data is clean.

Not every system needs the same oversight. Copilot: light governance, with usage policies and monitoring. The customer chatbot: policy-based, with escalation paths, retrieval grounding, and output review. Agentic systems on credit or compliance data: strong governance, with full audit trails, the Section 4 consensus architecture, and human authorization for decisions carrying regulatory exposure.

The EU AI Act makes this concrete. Consumer creditworthiness assessment and credit decisioning are likely high-risk, as are risk assessment and pricing in covered life and health insurance. Other underwriting is not automatically high-risk; classify it by purpose, impact, and applicable law. The EBA

consultation (Signal 2) directly affects our German operations, indirectly the shared infrastructure supporting the UK. The compliance timeline overlaps our eighteen-month window.

Proposal C separates beyond cost. If rules tighten, A and B renegotiate, migrate data, or rebuild integrations; with C, compliance becomes a module in an architecture we control. We add it. We do not rebuild. On IP: under A, Microsoft owns the models and we own the prompts; under B, the vendor owns the weights; under C, we own everything, including the compliance logic. Who owns the IP underpinning credit decisions is a fiduciary question, not a technical one.

Governance is the floor: not the headline, the foundation. We should leave this process motivated to govern AI well, not afraid to deploy it.

6. The Gates

We recommend staged capital release against three conditions. If any is not met, we pause and reassess.

Gate 1: Data readiness. At least three of five data systems must meet defined quality thresholds before any agentic workflow moves to production. The CDO's six-to-eight-month estimate is the pacing constraint. If the data is not ready, the system is not ready.

Gate 2: Governance operational. The working group must move the policy from draft to approved and operational, with risk tiering and a named accountable owner, before we scale beyond internal pilot. If the framework is not operational by Month 6, we do not expand.

Gate 3: Team in place. At least two senior AI engineers hired and onboarded before we commit capital beyond the foundation phase. We cannot execute Proposal C with the current team. If we cannot attract the right senior talent within six months, we reassess scope; we do not lower the hiring bar.

If the acquisition (Signal 3) is pursued, due diligence must clear independently before that capital is released: clean IP ownership, indemnification terms, key-employee retention. A failed acquisition does not stop the initiative. It changes the timeline.

7. What This Means for My Own Work

Two contexts of mine map directly to this case: a large enterprise infrastructure company where I led developer and platform initiatives across a field organization of several thousand, and a small marine tourism operation where I ship production AI systems as a fractional AI engineer.

The enterprise context mirrors Northbridge: structural vendor dependence, shadow IT in the field before leadership acknowledged it, no governance infrastructure, and a field organization, not the technology team, holding the keys to adoption. I learned there that developers are a trailing indicator: if the field does not adopt first, the ecosystem never materializes. That discipline is the logic behind this recommendation.

The small-company context is where I feel the Northbridge constraints personally. I built a production AI system for Sanctuary Cruises: it ingests live environmental data and one sentence from a non-technical crew member, then generates publication-quality naturalist reports published daily to a live site driving bookings, every report human-reviewed before it goes live. Strong answer-engine

optimization helped drive a 30% booking increase in eight weeks, and Sanctuary Cruises now appears as the leading recommendation for whale watching out of Moss Landing, backed by more than a thousand reviews and a 4.9 rating, letting a small operator compete with businesses three to four times our capacity.

I chose Gemini Flash as the intelligence layer because it is the best model for this use case today, but I built the abstraction layer Northbridge lacks: swapping vendors is a token change, not a rebuild. I built it deliberately: AI-sector volatility is unlike anything in my thirty years in technology, and the Copilot billing shift in Section 3 is the exact risk it protects me from.

The most useful lesson from this case: size governance to the stakes. Northbridge is a bank moving into credit decisioning: strong controls, audit trails, consensus validation. Sanctuary Cruises is a whale-watching boat publishing naturalist observations. The captain knows what he saw. Human review of every report is the right governance for the actual risk, not a gap to fill. Bank-grade oversight here would be the same mistake in reverse: control theater slowing the operation without reducing real exposure.

The Northbridge constraint that maps most closely to my current work is the category shift itself: moving from low-risk AI-generated content to systems that take actions involving customers, bookings, communications, and operations, which is exactly our next roadmap request: crew-facing agents for trip logging and booking communications. On Monday, I will define that threshold explicitly: report generation stays human-reviewed and light; anything acting autonomously on bookings or customer communication moves up a tier before it ships. And on Monday, I also hope we see orcas: orca sightings are our single biggest booking driver, and an automated Captain's Log and social media pipeline is what lets us turn a sighting into a bigger bet.

Decision Framework Annex

Northbridge Financial Group | AI Transformation Initiative

Supporting analysis for the Strategic AI Brief. This annex makes the evaluation visible: how each proposal scores across the dimensions that drive the recommendation of Proposal C.

Prepared by Luke Kilpatrick | University of Waterloo AI-CTO Program | July 2026

Strategy and Model Positioning

Evaluation Dimension	Proposal A Fast Follower (Microsoft/Azure)	Proposal B Configured Platform (Vendor)	Proposal C Custom Build (Recommended)
AI Adoption Spectrum	Configured. Extends existing Azure OpenAI integration; no owned models.	Configured. Licenses a pre-built financial-services platform on vendor-owned models.	Custom. Fine-tunes open foundation models on our data; fully owned platform.
Foundation Model Strategy	Closed API. Dependent on Microsoft's model releases, behavior, and pricing.	Vendor-hosted weights. Vendor owns the training pipeline; we own application outputs.	Open weights plus fine-tuning and retrieval grounding. We own weights and pipeline.
Vendor Lock-in Risk	HIGH. Single-vendor dependency on models, infrastructure, and pricing. EA expires in 14 months.	HIGH. Vendor owns weights and pipeline; switching means rebuilding compliance agents.	LOW. Own weights, data, and platform. Can change base models with re-tuning, not a rebuild.
Exit Flexibility	Weak. Prompts portable in theory, deeply coupled to Azure services in practice.	Weak to moderate. Outputs portable; compliance logic embedded in vendor platform.	Strong. All IP owned. Model, data, and architecture are organizational assets.

Frameworks: Week 11 foundation-model strategy (closed API vs. open weights vs. fine-tuning) | Week 5 vendor dependency

Risk, Capability, and Regulatory Exposure

Evaluation Dimension	Proposal A Fast Follower (Microsoft/Azure)	Proposal B Configured Platform (Vendor)	Proposal C Custom Build (Recommended)
Data Readiness Dependency	Moderate. Can scope to cleaner systems first; agentic workflows still gated by the 6-8 month remediation.	High. Pre-built agents assume clean, structured inputs. 3 of 5 systems have quality issues.	High. Fine-tuning requires clean data; quality issues degrade the model directly.
Capability: Required vs. Available	Lowest gap, but 31% Copilot use shows adoption, not deployment, is the real problem.	Moderate gap. Vendor handles model complexity; internal team configures and validates.	Largest gap. Needs senior hires. The 6-person ML team is the knowledge base, not the delivery team.
Contracting Risk (flexibility, indemnification)	High. EA renews in 14 months with a pricing review pending; limited flexibility clauses and AI-specific indemnification.	Moderate. Specialized vendor may negotiate, but smaller counterparty; indemnification for compliance outputs is critical.	Low for the platform (no core vendor contract). Acquisition contract must clear IP, indemnification, and retention terms.
Governance Overhead Required	Light to moderate. Inherits vendor controls, but we still tier and monitor usage; less control means less to build, less to verify.	Moderate. Configure and validate vendor guardrails; audit depends on vendor transparency.	High but owned. We build the full framework: risk tiering, audit trails, consensus, human authorization. Highest effort, strongest control.
EU AI Act / Sovereignty (Signal 2)	Exposed. Credit and covered-insurance uses likely high-risk; non-EU (Azure) hosting adds sovereignty risk.	Exposed. Same high-risk uses; non-EU vendor hosting adds the same sovereignty concern.	Materially reduced. We control deployment location and hosting; compliance added as a module.
Agentic Risk Fit (precision vs. complexity)	Pre-built agents limit control over guardrails and consensus design.	Vendor agents may not match our specific regulatory interpretation needs.	Full control to match architecture to each workflow: consensus and human authorization where stakes demand.

Frameworks: Week 10 risk categories | Week 12 agentic precision/complexity | Week 5 bias-variance (clean data)

Economics and Overall Assessment

Evaluation Dimension	Proposal A Fast Follower (Microsoft/Azure)	Proposal B Configured Platform (Vendor)	Proposal C Custom Build (Recommended)
Year 1 Cost (stated)	**$4.2M**	**$7.8M**	$11.4M stated Year 1 build cost. Acquisition capital, if approved, must fit inside the $10M execution bucket and clear diligence.
Cost Behavior at Scale	Volatile. Signal 1 pricing review of 30-50% plus consumption billing; Year 2 realistically $5.5-6.3M.	Moderate. Licensing plus maintenance; each fine-tuned version is another supply-chain item.	Most predictable over 24 months. We own compute, models, and data pipeline.
Time to First Production	3-4 months. Fastest first output.	5-6 months.	8-10 months; 6-8 for the compliance module with the acquisition.
Operationalization Velocity (production to governed, scaled use)	Fast to first output, but governance and audit depend on vendor pace; scaling is gated by pricing and sovereignty exposure.	Moderate. Pre-built modules speed rollout, but validating vendor guardrails against our rules slows trusted scale.	Slower to start, fastest to safely scale once built: owned governance and controls let us extend to new workflows without renegotiation.
Strategic Signal Exposure	S1: directly exposed. S2: exposed. S3: not usable.	S1: indirect. S2: exposed. S3: not usable.	S1: reduced. S2: reduced. S3: directly usable.
Overall Assessment	The fastest path to a visible win and the right choice for a scoped, low-risk internal use case. But single-vendor pricing volatility and non-EU sovereignty exposure make it the wrong long-term home for credit and compliance workloads. Best used as a near-term complement, not the platform of record.	A middle ground that offers neither A's speed nor C's ownership. Pre-built compliance modules are attractive, but the vendor owns the weights on our most critical workflows, which is a structural lock-in risk for a regulated institution. Reasonable only if building and owning is truly off the table.	**The strongest long-term position on control, exit flexibility, predictable 24-month economics, sovereignty, and owned compliance logic. It costs the most upfront and takes longest to first output, so it must be sequenced behind governance, data, and hiring gates. The acquisition accelerates it but is not required. Recommended.**

Capital plan: $3M foundation | up to $10M acquisition + platform execution | $2M gated reserve. Total within the $15M envelope.

Sequencing and Gate Conditions

Proposal C is recommended, but not deployed on Day 1. Capital releases in stages against three gates.

Months 1-6: Foundation

$3M released now. Stand up governance, begin data remediation, hire two to three senior engineers. The ML team interviews and trains alongside the CTO.

Months 4-12: Accelerate

Acquisition due diligence (Signal 3). If it clears on IP, indemnification, and retention, it takes 6-8 months off the compliance timeline. Not load-bearing.

Months 8-18: Own & Scale

Platform to production for high-stakes workflows. EU AI Act compliance added as a module. First production capability in 8-10 months.

Gate	Condition	Stop-or-Pivot Trigger
1. Data readiness	At least 3 of 5 data systems meet quality thresholds before any agentic workflow reaches production.	Data not ready by remediation milestone means the system is not ready.
2. Governance operational	Policy moves from draft to approved and operational, with risk tiering and a named owner, before scaling beyond pilot.	Not operational by Month 6 means we do not expand.
3. Team in place	At least 2 senior AI engineers hired and onboarded before capital commits beyond the foundation phase.	No senior talent in 6 months means reassess scope, not lower the bar.

$12M of the $15M envelope stays behind these gates. A failed acquisition changes the timeline, not the initiative.

BOARD RECOMMENDATION MEMO

AI Transformation Initiative

To:	Board of Directors, Northbridge Financial Group
From:	Luke Kilpatrick, AI Transformation Lead
Date:	July 2026
Re:	Recommendation on the $15M AI transformation decision

RECOMMENDATION

We recommend Proposal C: build and own the AI platform, on open foundation models we fine-tune and control, grounded in our own data pipeline. It is not the cheapest or fastest option, but it is the only one that lets Northbridge change models, infrastructure, and compliance logic as the regulatory and pricing environment shifts. It fits exactly within the $15M envelope over 18 months, released in stages against the conditions below.

WHY THIS PATH

- **Strategic control and exit flexibility.** A and B lock us into a single vendor's architecture, pricing, and model roadmap. Proposal C means we own the compute, models, data pipeline, and compliance logic, and can adopt best-of-breed systems as the market moves.
- **Predictable economics as usage scales.** The recent GitHub Copilot move to token-based billing shows how vendor AI costs turn variable and rise with adoption. Proposal A's $4.2M Year 1 ignores the announced 30 to 50 percent Microsoft pricing review and our EA expiry in fourteen months.
- **Regulatory defensibility and sovereignty.** Credit decisioning and covered insurance pricing are likely high-risk under the EU AI Act, and the EBA consultation raises scrutiny of non-EU infrastructure. With Proposal C, EU compliance becomes a module in an architecture we control, and we own the IP underpinning every credit decision.

TOP THREE RISKS WE ARE ACCEPTING

- **Execution risk.** Proposal C depends on hiring senior AI talent in a tight market. If we cannot hire within six months, scope must be reassessed rather than the bar lowered.
- **Data and governance risk.** The platform cannot safely scale until data remediation and operational governance are real, not draft artifacts. A confidently wrong model on bad data is worse than none.
- **Acquisition risk.** The acquisition may fail diligence or distract execution. It accelerates the plan but is not required, which is why it sits behind independent due diligence.

CONDITIONS FOR SUCCESS

- **Data readiness.** At least three of five data systems meet defined quality thresholds before any agentic workflow reaches production. The CDO's six-to-eight-month remediation estimate is the pacing constraint.
- **Governance operational.** AI governance policy moves from draft to approved and operational, with risk tiering and a named accountable owner, before we scale beyond internal pilot.
- **Team in place.** At least two senior AI engineers hired and onboarded before capital is committed beyond the foundation phase. The existing six-person ML team is the knowledge base, not the delivery team.

DECISION REQUESTED

We ask the board to approve Proposal C, release the $3M foundation-phase capital now, authorize recruitment of at least two senior AI engineers, permit acquisition due diligence, and hold the remaining $12M behind the three gates above.