MatchPoint Research Working Paper No. 2026/T03

Generative AI for Family-Office Operations: From Reporting to Portfolio Intelligence

A governed architecture for reporting, document intelligence and accountable portfolio decision support

Chennakeshav (CK) Adya
Independent Researcher
July 2026

ABSTRACT

Background. Family offices combine institutional investment information with private-company, entity and personal records. Generative AI aligns with reporting, document review and information-retrieval work, while privacy, data quality and transaction risk constrain deployment. **Objective.** This paper examines current family-office evidence and develops a governed route from reporting support to portfolio intelligence. **Approach.** The analysis reviews 17 primary and authoritative sources, including global family-office surveys, regional regulatory evidence, private-markets reporting guidance, accounting standards and AI risk frameworks. Findings are classified as observed family-office evidence, authoritative guidance, Matchpoint synthesis or Matchpoint scenario analysis. **Findings.** Citi reports that 22% of surveyed family offices use AI for operational tasks or investment analysis and 16% use it for investment-performance reporting. Reporting, account reconciliation and investment intelligence appear as current use areas. The principal design requirement is separation: reconciled systems retain official records; deterministic engines calculate performance, exposure, liquidity and valuation outputs; controlled retrieval supplies approved evidence; the model drafts cited explanations; accountable people validate material outputs and authorise actions. **Implications.** A family office can begin with a bounded read-only use case, a fixed evaluation set and explicit security, quality and value gates. Portfolio intelligence requires controlled metadata, lineage, permissions, model monitoring and transaction-plane separation.

Highlights

- Operational intelligence is the present entry point for family-office GenAI.
- Portfolio answers require reconciled data, dates, definitions and source lineage.
- Deterministic systems retain calculation, valuation and transaction authority.
- Controlled retrieval, least privilege, testing and human approval are architectural controls.
- No source in the evidence register establishes family-office investment alpha from GenAI.

JEL Classification: G11, G23, G32, M15, O33

Keywords: generative AI, family office, portfolio intelligence, investment reporting, private markets, retrieval-augmented generation, data governance, AI risk management, human oversight

Research paper only. This document is not legal, tax, accounting, cybersecurity or investment advice. Matchpoint scenario analysis is illustrative and is not observed productivity, a forecast or evidence that GenAI will improve investment returns.

1. INTRODUCTION

Family offices operate at the intersection of institutional investment management, private-company ownership, household administration and multigenerational governance. Their information arrives through custodians, banks, fund administrators, private-market managers, operating companies, advisers, tax specialists and family members. It spans structured files, portals, emails, investment memoranda, capital-call notices, partnership accounts, board papers, legal documents and personal records. The operating challenge is therefore an information-control problem before it becomes an artificial-intelligence problem.

Generative artificial intelligence can search, summarise, classify and draft across large bodies of text. These capabilities align with several activities that consume family-office time: consolidating reports, reviewing manager communications, extracting terms from private-market documents, preparing committee papers and answering questions about portfolio exposures. Recent family-office research indicates that adoption is moving in this direction. Citi reports that 22% of surveyed family offices used AI for operational tasks or investment analysis and 16% used it for investment-performance reporting [1,2]. Deloitte's earlier survey of 354 single family offices found 12% using AI-driven solutions, alongside materially broader use of data analytics [3]. UBS respondents identified financial reporting or data visualisation and text analysis as the most likely operational AI uses over the following five years [4].

The same operating environment makes uncontrolled deployment particularly hazardous. A family office may hold intimate personal data, entity structures, tax information, unreleased transactions, private-company records, bank details and confidential manager material. It also depends on valuations and cash-flow records that differ in source, date, currency, methodology and degree of verification. A fluent answer assembled from the wrong reporting period can appear credible while being decision-useless. A summarisation tool connected to email or document stores can expose information beyond an employee's authority. An agent with payment access can turn a model error or malicious instruction into a financial event.

This paper develops a governed route from reporting automation to portfolio intelligence. Its core proposition is that the generative model should operate above reconciled systems of record and deterministic analytical engines. The model can retrieve approved evidence, explain calculations, assemble commentary and prepare drafts. Accounting records, valuations, exposure calculations, transaction instructions and approval authority remain within controlled systems and accountable human processes.

The paper asks five questions. First, what does current evidence establish about family-office AI adoption? Second, which use cases have sufficient data and control readiness to justify deployment? Third, what reference architecture separates records, calculations, retrieval and generation? Fourth, which risks require controls before sensitive information enters an AI workflow? Fifth, how can a family office test value through bounded pilots and measurable gates?

Four evidence labels are used. **Observed family-office evidence** refers to survey or interview findings reported by the cited provider. **Authoritative risk or reporting guidance** refers to standards, laws and institutional frameworks. **Matchpoint synthesis** identifies an architecture, control or operating-model recommendation derived from the evidence. **Matchpoint scenario analysis** identifies an illustrative timing, staffing, cost or value case; it is not observed performance or a forecast.

Survey findings describe their cited samples and fieldwork periods. They do not establish causation, implementation quality or investment performance. Matchpoint frameworks require adaptation to each office's systems, legal obligations, risk appetite and governance.

1.1 Contribution and scope

The paper contributes a use-case ladder, a governed data model, a reference architecture, a risk-control map, a buy-build-outsource matrix and a phased implementation roadmap. The framework is designed for single family offices and for investment organisations supporting family capital. It addresses public and private assets, while recognising that private-market reporting creates particular data-timing and document-processing challenges.

The analysis does not recommend securities, managers, allocations or technology vendors. It does not represent generated commentary as accounting, valuation, legal or investment advice. Any illustrative operating scenario is explicitly labelled Matchpoint scenario analysis. The paper uses the UAE Personal Data Protection Law overview as a relevant jurisdictional prompt for UAE-based offices; legal applicability, free-zone regimes and cross-border duties require qualified counsel [14].

1.2 Executive conclusion

Five conclusions follow from the evidence.

First, operational intelligence is the current entry point. Reporting, account reconciliation, document review and information retrieval appear repeatedly in family-office and financial-sector evidence [1,4,12,17]. These uses can be bounded by source collections, reporting periods, approval rules and measurable quality tests.

Second, data governance determines answer quality. Portfolio intelligence requires records with defined owners, as-of dates, valuation dates, currencies, entity mappings and verification status. A model connected directly to inconsistent documents reproduces those inconsistencies in fluent form.

Third, deterministic calculation and generative explanation require separate components. Exposure, performance, cash, tax lots, commitments and valuations should be calculated by approved systems or controlled code. The generative layer can cite and explain those outputs. It should not create the official number.

Fourth, access control and human approval are architectural features. NIST identifies confabulation, data privacy, information security, human-AI configuration and value-chain integration among GenAI risks [10]. NCSC and CISA guidance emphasises threat modelling, supply-chain security, least privilege, logging and incident management [11]. These controls belong in the product design and operating process.

Fifth, a staged programme should begin with low-authority workflows. A pilot can retrieve and summarise an approved document set, measure answer support and review effort, and remain disconnected from transaction release. Higher-materiality use cases proceed only after data, security, validation and governance gates have been passed.

2. WHAT THE EVIDENCE SAYS ABOUT ADOPTION

2.1 Family-office technology foundations

Deloitte's 2024 family-office study provides a useful pre-GenAI operating baseline. The survey covered 354 single family offices and was conducted from September to December 2023 [3]. It reported that 43% were developing or rolling out a technology strategy. Cloud applications or services were used by 87%, and identity and access management by 61%. Data analytics had broader penetration than AI: 55% used analytics to a moderate or large extent in investments and 42% did so in operations. AI-driven solutions were reported by 12% [3].

These findings show that AI deployment sits within an existing technology stack. Cloud infrastructure, identity systems, portfolio platforms, accounting tools and analytical datasets shape what an office can implement. They also create dependencies that require ownership and monitoring. The Matchpoint synthesis is that an AI programme should begin with a data and system inventory rather than with a model demonstration.

Citi's 2025 survey reports subsequent movement. More than one fifth of respondents, 22%, had automated some operational tasks or were conducting investment analysis or forecasts with AI, compared with approximately 13% in 2024 [2]. AI use in investment-performance reporting reached 16%, while portfolio construction reached 13% [2]. These are self-reported use measures. They do not show the proportion of a workflow automated, the quality of outputs or the strength of controls.

2.2 Adoption barriers

Citi's 2025 report identifies the internal capability gap as the leading barrier. Lack of internal expertise was reported by 57% of respondents, lack of awareness of available options by 34%, cybersecurity or privacy concerns by 28%, and uncertain return on investment by 25%. Only 16% reported no technology-adoption barrier [2]. The order matters. A governance programme that addresses privacy without developing process knowledge may remain unable to select, configure or validate useful systems.

Citi's 2026 interviews describe varied adoption, from offices with company-wide use expectations to offices still resolving privacy and security concerns [1]. The report identifies reporting, account reconciliation and investment intelligence as commonly mentioned use areas [1]. It also records demand for email summarisation, meeting transcription, document review and reporting automation. These uses share two characteristics: they involve language-intensive work, and an experienced user can compare the draft with source material.

UBS's 2025 survey provides a forward-looking complement. Respondents were most likely to identify financial reporting or data visualisation and text analysis as AI uses within operations over the following five years [4]. An expectation is not a completed deployment. It does reinforce the direction observed in Citi's reporting and interview evidence.

2.3 Regional financial-services context

The DFSA's 2025 survey captured responses from 661 Authorised Firms across banking, capital markets, wealth and asset management, and fintech; participation was 88% [6]. It reported AI use by 52% of firms. Sixty per cent had some form of governance structure, while 21% lacked clear accountability or oversight mechanisms [6]. This is a DIFC financial-services sample and is not family-office-specific.

The regional evidence has two implications for family offices. First, external managers, banks and service providers are likely to introduce AI into their own processes at different speeds. Vendor diligence therefore needs questions about upstream AI use, data handling and accountability. Second, local financial-services practice is developing while governance maturity remains uneven. A family office cannot assume that provider status alone establishes adequate AI controls.

2.4 Interpreting the evidence

Source	Sample or method	Observed result used in this paper	Interpretation boundary
Deloitte 2024 [3]	354 single family offices; fieldwork September-December 2023	12% AI use; 55% investment analytics; 42% operational analytics	Earlier adoption baseline; self-reported use
Citi 2025 [2]	Global family-office survey	22% operational or investment-analysis AI use; 16% performance-reporting use	No measure of workflow depth or output quality
Citi 2026 [1]	Interviews with principals and CIOs across four regions plus prior survey evidence	Reporting, reconciliation and investment intelligence are common use areas	Qualitative interviews; no universal operating model
UBS 2025 [4]	Global family-office survey	Reporting or data visualisation and text analysis lead expected operational uses	Future intention; no measured benefit
DFSA 2025 [6]	661 Authorised Firms; 88% participation	52% AI use; 21% lacked clear accountability or oversight	Financial-services context; not family-office-only

The evidence supports careful adoption. It does not establish GenAI-driven family-office alpha. No source in this register demonstrates superior realised investment returns caused by a family-office GenAI system. The immediate evidence concerns operations, information processing and emerging governance.

3. THE USE-CASE LADDER

3.1 Four levels of authority

The Matchpoint use-case ladder classifies workflows by the authority granted to the system.

Level 1: productivity support. The model drafts, translates, transcribes, classifies or summarises material supplied by a user. It has no standing access to systems of record and no ability to change data. Examples include preparing a first draft from an approved investment memorandum or summarising a selected meeting transcript.

Level 2: controlled reporting support. The model retrieves from an approved document collection or data mart, produces cited commentary and flags missing or inconsistent fields. It remains read-only. Examples include drafting a quarterly narrative from validated performance tables or comparing manager letters across periods.

Level 3: portfolio intelligence. The model answers cross-source questions, connects documents to deterministic calculations and supports scenario discussion. It can assemble committee materials and generate exception explanations. Every answer exposes source, as-of date and data status. Material outputs require named reviewers.

Level 4: controlled workflow orchestration. The system creates tasks, requests information, prepares instructions or routes approvals across systems. Transaction release remains outside the generative model. Any move from draft to action follows deterministic rules, segregation of duties and authorised human approval.

Authority increases from Level 1 to Level 4. Data sensitivity, validation burden, monitoring and incident impact increase with it. An office should approve use cases individually rather than approve a model for unrestricted use.

3.2 Use-case screen

Each candidate use case should be scored across six dimensions.

- **Decision materiality:** What financial, legal, tax, privacy or reputational consequence could follow from an error?
- **Data readiness:** Are the sources complete, current, reconciled, permissioned and identifiable by date and owner?
- **Answer verifiability:** Can a reviewer check the output against cited evidence or a deterministic calculation?
- **Process frequency:** Does the workflow occur often enough to justify design, testing and monitoring?
- **Authority required:** Is the system read-only, able to create a draft, able to change a record, or able to initiate an action?
- **Control coverage:** Are access, retention, logging, testing, review and incident processes already defined?

The highest-priority pilots combine frequent work, ready data, low authority and high verifiability. A document-search assistant over approved investment memoranda generally scores more favourably than an agent that can initiate payments. The comparison reflects control burden rather than excitement or model capability.

Use case	Typical authority	Principal data dependency	Validation method	Initial readiness
Meeting transcript and action draft	Draft only	Approved recording and participant list	Human comparison with transcript	Higher
Manager-letter summarisation	Read-only retrieval	Versioned letters and fund mapping	Citation check and exception sample	Higher

Use case	Typical authority	Principal data dependency	Validation method	Initial readiness
Quarterly commentary draft	Read-only retrieval plus deterministic data	Reconciled performance and exposure tables	Number lock and reviewer sign-off	Medium
Capital-call notice extraction	Read-only extraction	Original notice and fund master	Field-level reconciliation	Medium
Portfolio question answering	Read-only retrieval	Cross-asset data model and dated calculations	Evidence completeness and calculation trace	Medium
Liquidity scenario narrative	Decision support	Approved scenario engine and cash-flow data	Independent scenario rerun	Medium to lower
Payment instruction preparation	Draft plus workflow	Bank details, approvals and fraud controls	Dual authorisation and independent verification	Lower
Autonomous trading or payment release	Execution authority	Multiple critical systems	Material control burden	Excluded from the proposed initial programme

3.3 Value should be measured as process evidence

Productivity claims require a defined baseline. A pilot should measure elapsed time, reviewer time, exception rate, evidence coverage, rework and user acceptance before and after the controlled intervention. The baseline needs the same document set, service level and output specification. Time saved on the first draft may be offset by source checking or correction. Measurement therefore covers the complete process.

Quality measures should include factual support, numerical fidelity, citation correctness, omission rate and escalation behaviour. User satisfaction is useful and insufficient on its own. A fluent response can receive a high user score despite stale or incomplete evidence. Decision-grade evaluation requires objective checks alongside user judgement.

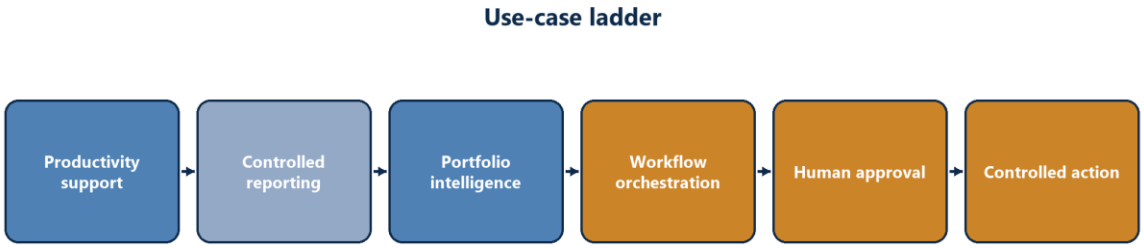

Figure 1. The use-case ladder from productivity support to controlled workflow orchestration
Source: Matchpoint framework, informed by the family-office use cases in [1-4].

4. THE GOVERNED DATA FOUNDATION

4.1 Systems of record and systems of intelligence

A system of record holds the official position, transaction, entity, document or approval state. A system of intelligence combines approved records and analysis to support questions and decisions. A generative model belongs in the intelligence layer. It should not silently become a second ledger.

The distinction is especially important in private markets. A quarterly report may arrive weeks after the period end. Capital-account statements, cash-flow notices and valuation files may have different dates. A general partner can restate a value. A family office can maintain its own cash-flow record while relying on a manager

or administrator for reported net asset value. An answer about current exposure therefore needs both the source date and the office's policy for carrying values forward.

ILPA's Reporting Template v2.0 promotes more uniform reporting for fees, expenses and carried interest [7]. The template offers a relevant example of a canonical external schema. It does not cover the entire family-office portfolio, and it does not validate received data. The Matchpoint synthesis is to map each external report into controlled internal fields, preserve the source document, run validation rules and expose exceptions before information reaches the generative layer.

4.2 Minimum metadata

Every retrievable investment record should carry metadata that allows a user and the system to interpret it. The required fields vary by asset class. A common minimum includes:

- legal entity and beneficial ownership context;
- asset, fund, account or manager identifier;
- source system and source document;
- reporting period, as-of date and valuation date;
- currency, unit and foreign-exchange treatment;
- data owner and approving function;
- verification or reconciliation status;
- confidentiality classification and permitted audience;
- version and superseded-record link;
- retention requirement and deletion status.

An answer should expose the fields that determine meaning. A statement such as “private-equity exposure is X” is incomplete unless X has a defined entity perimeter, currency, date, valuation basis and treatment of unfunded commitments. The interface can make those choices visible through answer headers, footnotes or expandable evidence cards.

4.3 Data-quality controls

Data-quality rules should run before retrieval. They can test required fields, date relationships, currency consistency, duplicate records, unexplained changes, missing valuations, stale reports and mapping breaks. Deterministic reconciliations can compare cash activity with bank or custodian records, capital calls with partnership cash flows, and portfolio totals with accounting control accounts.

The model can assist with exception triage. It can group similar breaks, explain which source fields disagree and draft a request to the responsible provider. It should not close the exception or overwrite the official record without the defined review process.

Data lineage connects each derived output to its sources and transformations. A portfolio-intelligence answer needs lineage across the retrieved passage, calculation output and narrative. When a source changes, affected answers and reports should be identifiable. This supports correction, incident review and repeatability.

4.4 Valuation boundary

IFRS 13 defines fair value as an exit price in an orderly transaction between market participants at the measurement date and provides a measurement framework where another IFRS requires or permits fair value [8]. Its relevance here is the discipline of a defined measurement process, assumptions and disclosures. Applicability depends on the reporting entity and accounting framework.

The Matchpoint control principle is that an LLM should not create the official valuation. It can retrieve the approved value, explain the valuation method recorded by the office, compare reporting dates and identify

missing support. Valuation models, accounting policy, approved adjustments and professional judgement remain within governed processes. Generated explanations should cite the applicable value, date and source.

Data object	Official owner	Deterministic control	Permitted GenAI role
Cash balance	Treasury or accounting system	Bank reconciliation	Explain movement and retrieve supporting items
Public-security position	Custodian or portfolio record	Position and price validation	Summarise exposure and relevant research
Private-fund NAV	Approved manager or administrator record	Date, currency and restatement checks	Retrieve report, explain change and flag staleness
Unfunded commitment	Controlled commitment ledger	Notice and cash-flow reconciliation	Answer liquidity questions using approved calculation
Performance return	Approved performance engine	Methodology and independent calculation	Draft commentary around locked figures
Entity ownership	Legal or corporate-secretarial record	Authorised update and version control	Retrieve approved structure and document links
Payment instruction	Treasury workflow	Verified beneficiary and dual approval	Draft request or checklist only

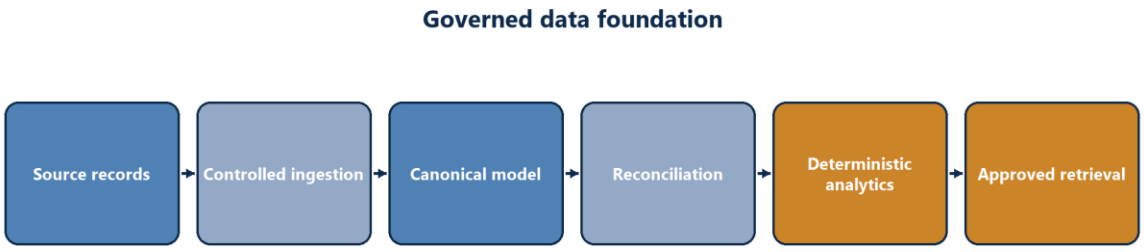

Figure 2. The governed data foundation: records, validation, lineage and approved retrieval
Source: Matchpoint framework, informed by [7-12].

5. REFERENCE ARCHITECTURE

5.1 Six layers

The Matchpoint reference architecture contains six layers.

Layer 1: source systems. Custodians, banks, accounting systems, portfolio platforms, customer-relationship systems, document stores, data rooms, email and external portals remain the points of origin.

Layer 2: controlled ingestion. Connectors capture files and records, retain source identity, classify sensitivity, scan content and map fields. Ingestion should quarantine invalid or unauthorised material. It should preserve an immutable copy or reference where policy requires it.

Layer 3: reconciled data and document catalogue. A canonical model aligns entities, accounts, assets, funds, managers, currencies and dates. Data-quality rules create exceptions. The catalogue records permissions, versions and provenance.

Layer 4: deterministic analytics. Approved engines calculate returns, exposures, liquidity, commitments, concentration, scenarios and accounting outputs. Their inputs, methods and versions are controlled. Results can be queried by the intelligence layer.

Layer 5: controlled retrieval and generation. The system retrieves only material the requesting user is entitled to see. It supplies relevant passages, metadata and deterministic outputs to the model. The model creates an answer or draft with citations, dates and uncertainty or exception indicators.

Layer 6: review and action. A named reviewer accepts, corrects, rejects or escalates the output. Any downstream record change or transaction follows a separate workflow with deterministic controls and authorised approvals.

5.2 Retrieval design

Retrieval-augmented generation narrows the information supplied to the model. It can improve evidence access and citation. It does not eliminate confabulation, stale sources, malicious instructions or retrieval errors. NIST's GenAI profile treats information integrity, privacy, security and human-AI configuration as distinct risk areas [10]. Each needs its own controls.

A document should be segmented in a way that retains headings, tables, footnotes and surrounding context. Metadata filters should apply before semantic similarity. A user asking about one family entity should not retrieve records for another entity solely because the language is similar. Date and document-status filters should prevent superseded reports from being treated as current. Where multiple sources disagree, the system should present the disagreement and route it as an exception.

The answer policy should require evidence for every material factual statement. Evidence coverage can be measured as the proportion of claims supported by an accessible citation. A citation is useful only when it points to the correct source passage or deterministic output. Evaluation therefore checks citation presence and citation entailment.

5.3 Numerical controls

Numbers should enter the prompt as locked outputs with names, units, dates and calculation identifiers. The model may format or discuss them. The publication workflow should compare every number in the generated narrative with the approved data payload. A mismatch should block release and create an exception.

For a quarterly portfolio report, the process can operate as follows:

- the performance engine calculates approved returns;
- the exposure engine calculates weights and concentrations;
- the liquidity engine calculates commitments and cash-flow scenarios;
- the reporting service creates a signed data payload;
- retrieval supplies approved manager commentary and market context;
- the model drafts narrative linked to the payload and evidence;
- automated checks compare narrative numbers with the payload;
- portfolio, finance and compliance reviewers approve their respective sections;
- the final report is stored with the prompt, evidence package, model version and approvals.

This approach lets the model explain a result while preserving the calculation authority elsewhere.

5.4 Access and isolation

Permissions should be inherited from authoritative identity and document systems. Role and attribute rules can incorporate family branch, legal entity, geography, confidentiality and project membership. Least privilege means that the retrieval service sees only what the approved use case needs and the user is entitled to access [11].

High-sensitivity collections can require dedicated environments, private networking, restricted retention, customer-managed encryption keys or on-premises components. The choice depends on threat model, vendor architecture, legal requirements and operating capacity. The office should document where prompts, retrieved

content, outputs, logs and embeddings are stored; who can access them; whether they are used for model training; and how deletion propagates.

5.5 Logging and observability

The system should record user, use case, data sources, retrieval results, model and version, prompt template, output, automated test results, reviewer decisions and downstream actions. Logs require their own access and retention controls because they may contain sensitive content.

Operational monitoring should cover availability, retrieval failure, permission denial, latency, unsupported-answer rate, exception volume and provider changes. Quality monitoring should use a fixed evaluation set and sampled production reviews. Security monitoring should include unusual retrieval patterns, attempted instruction override, bulk extraction, data exfiltration indicators and anomalous action requests.

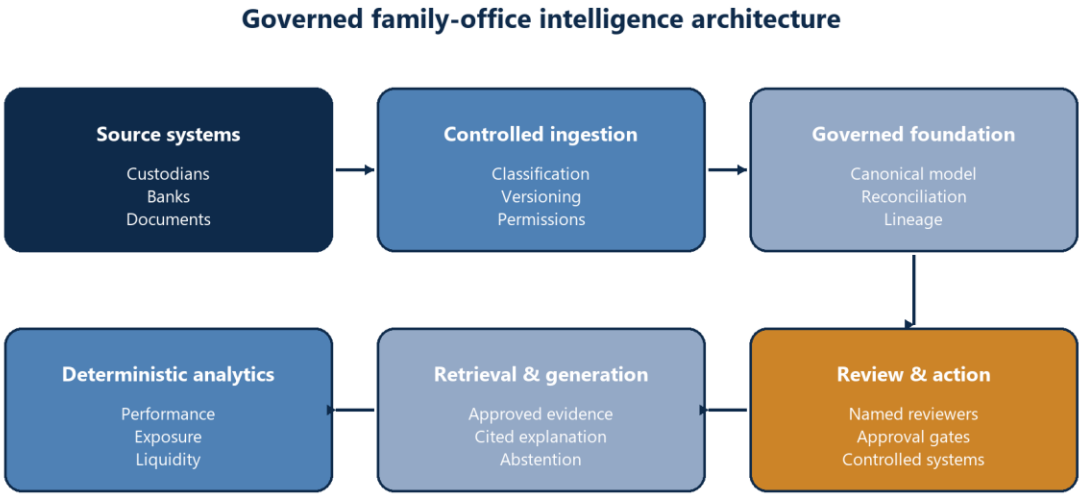

Figure 3. Reference architecture: systems of record, deterministic analytics, controlled retrieval, generation and approval
Source: Matchpoint architecture; governance and security controls informed by [9-13].

6. REPORTING AND DOCUMENT INTELLIGENCE

6.1 Reporting assembly

Family-office reporting often combines custodian files, administrator statements, manager reports and internal schedules. Citi's interviews describe demand for near-real-time reporting and note that monthly adviser reports can still be consolidated manually because of privacy and accuracy concerns [1]. The problem contains both structured and narrative work.

Automation should separate the two. Structured ingestion and deterministic calculations produce validated tables. GenAI can assemble commentary, explain material changes, summarise manager updates and flag missing support. A report should carry an explicit as-of date and data-completeness statement. Private-asset values with older dates should be identifiable rather than presented as contemporaneous market observations.

A controlled narrative template can define permitted sections, tone, evidence requirements and escalation rules. The model can state that a movement reflects contributions, distributions, market change, currency movement or reclassification only when the underlying calculation supplies that attribution. Unsupported explanations should be omitted and flagged for reviewer input.

6.2 Private-markets document extraction

Capital-call notices, distribution notices, quarterly reports, audited accounts, subscription documents and side letters contain recurring fields. GenAI and document-processing models can assist with extraction. The target output should be a proposed structured record linked to the exact document page and passage. Field-level confidence and validation rules guide human review.

For a capital call, relevant fields can include fund, investor entity, notice date, due date, amount, currency, purpose, bank details and contact. The system should compare beneficiary details with an independently maintained approved record. A change in bank details should trigger a fraud-control workflow. The model should never determine that a changed beneficiary is safe.

ILPA templates can support field standardisation for private-equity reporting [7]. External documents will still vary. The ingestion layer needs versioned parsers, exception handling and source retention. A model upgrade should be tested against a representative historical set before production use.

6.3 Manager monitoring

The intelligence layer can compare manager letters over time, identify changes in language, extract portfolio-company developments and map statements to existing risk themes. The output can support an analyst's review. It should preserve the manager's original wording through citations and distinguish manager-reported facts from the office's assessment.

Questions can be structured around exposure, valuation, operating performance, leverage, exits, fundraising, key-person events, litigation, regulatory matters and ESG or sustainability commitments where relevant. A generated red-flag list is a screening output. An experienced reviewer determines materiality and follow-up.

6.4 Committee packs

Committee-paper preparation combines data, prior decisions, manager communications and new recommendations. GenAI can retrieve previous minutes, assemble background and draft a decision record. The system should distinguish approved facts, management estimates, external claims and recommendations.

A decision page can include:

- question presented;
- entity and portfolio perimeter;
- data and valuation dates;
- verified facts with citations;
- unresolved exceptions;
- deterministic scenario outputs;
- recommendation owner;
- conflicts and required approvals;
- decision and follow-up actions.

The structure improves traceability. It does not delegate judgement to the model.

7. PORTFOLIO INTELLIGENCE

7.1 From search to questions

Document search answers “where is the information?” Portfolio intelligence answers “what does the approved information imply for this portfolio question?” The second task requires entity mapping, calculation services and a controlled ontology.

A question such as “What is our exposure to European logistics real estate?” may require public securities, direct holdings, private funds, debt positions, co-investments, operating companies and foreign-exchange effects. It also requires a definition of Europe, logistics and economic exposure. The system should retrieve the approved taxonomy and expose inclusions and exclusions. A single keyword search across documents cannot produce a decision-grade answer.

The architecture should convert the question into an explicit query plan. It identifies the requesting user, legal-entity scope, reporting date, asset classifications, calculations and supporting documents. Deterministic services produce the number. The generative layer explains the perimeter, major contributors, source dates and exceptions.

7.2 Private-market look-through

Look-through exposure is limited by available manager reporting. A fund may provide portfolio-company detail quarterly, semi-annually or only through a portal. The office should record the most recent look-through date, source and completeness. Estimated or mapped exposures should be labelled separately from manager-reported values.

GenAI can extract sector, geography and risk descriptions from manager materials. The resulting tags need review, especially where portfolio companies have several activities or regions. The office can maintain a controlled taxonomy and confidence status. The final exposure table should separate verified structured data, reviewed extracted data and management estimates.

7.3 Liquidity intelligence

Family-office liquidity connects bank balances, public assets, expected income, operating-company needs, capital calls, commitments, debt service, tax and planned distributions. A deterministic cash-flow engine should calculate scenarios. GenAI can retrieve assumptions, explain drivers and prepare commentary.

The system should distinguish contractual cash flows, manager forecasts, family plans and modelled scenarios. Dates and currencies require explicit treatment. An answer should disclose missing capital-call forecasts or stale commitment data. It should avoid presenting a point estimate as certainty.

7.4 Investment research and manager diligence

The model can organise research, compare documents and identify evidence gaps. It can create a first-pass manager profile from approved sources, extract stated strategy and terms, compare them with the office's mandate, and assemble diligence questions. External statements remain attributed to their source.

Final manager assessment requires human judgement, reference checks, track-record verification, operational due diligence, legal review and investment-committee governance. GenAI can support evidence handling and workflow completeness. The evidence register contains no basis for a claim that it selects superior managers or produces alpha.

7.5 Scenario commentary

Scenario engines can estimate portfolio effects under approved shocks or cash-flow assumptions. The model can translate those outputs into clear language, compare scenarios and identify concentrations. It should receive the result as a locked payload and cite the scenario definition. The scenario itself remains a model output with assumptions and limitations.

Intelligence question	Required deterministic component	Generative contribution	Required disclosure
What changed in performance?	Contribution and attribution engine	Draft explanation using cited manager and market material	Period, method and residual
Where is a risk exposure?	Position, taxonomy and look-through query	Explain perimeter and major contributors	Data dates and estimated mappings

Intelligence question	Required deterministic component	Generative contribution	Required disclosure
Can commitments be funded?	Cash-flow and liquidity scenario engine	Summarise drivers and exceptions	Scenario assumptions and missing forecasts
What did a manager report?	None beyond document controls	Retrieve, compare and summarise	Manager-reported status and source date
What action is recommended?	Approved analysis and governance workflow	Draft decision paper	Human owner, conflicts and approvals

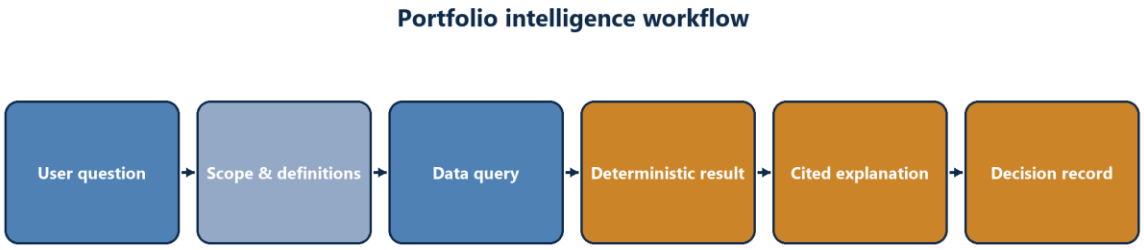

Figure 4. Portfolio intelligence: question, query plan, deterministic result, cited explanation and decision record
Source: Matchpoint framework.

8. DETERMINISTIC AND HUMAN-CONTROLLED BOUNDARIES

8.1 Calculations

Performance, exposure, valuation, cash, commitment and accounting calculations should execute in approved systems or controlled code. These processes require defined inputs, methodologies, versioning, test cases and reconciliations. The model can call a calculation service and explain the result. It should not recreate an official calculation from prose or infer missing numbers.

The narrative output should pass a number-lock test. Every amount, percentage, date and count is matched with the approved payload or cited source. A mismatch blocks publication. This control is mechanical and repeatable.

8.2 Investment judgement

Citi's 2026 report describes family offices using AI as an assistant while final investment decisions remain with experienced people [1]. This operating boundary aligns with broader governance principles emphasising human oversight [13,16].

Human review should be allocated by expertise. Portfolio professionals validate investment interpretation. Finance validates accounting and performance. Treasury validates liquidity and payment information. Legal and compliance review their respective obligations. Technology and security validate system operation and incidents. A generic approval button does not substitute for named accountability.

8.3 Transactions and payments

Transaction authority deserves a separate control plane. A GenAI interface may draft an instruction, populate a checklist or open a task. Beneficiary validation, mandate checks, limit checks, segregation of duties, multi-factor authentication and release occur through controlled systems.

Prompt content, retrieved documents and generated text should have no direct path to payment execution. The transaction system accepts only defined fields through an authorised workflow. Material changes, especially beneficiary or bank-detail changes, require independent verification. The architecture limits the consequence of a malicious document or erroneous model output.

8.4 Records and retention

Generated outputs become records when used in decisions, reports or communications. The office should define which prompts, sources, outputs, evaluations and approvals are retained. Retention must align with legal, regulatory, tax, employment, privacy and governance requirements applicable to the office.

Deletion is equally important. Removing a source document from the main repository may leave copies in embeddings, caches, logs, evaluation sets or vendor systems. Vendor diligence and architecture documentation should describe deletion propagation and exceptions.

9. RISK TAXONOMY AND CONTROL MAP

9.1 Confabulation and unsupported claims

NIST identifies confabulation as a GenAI risk [10]. In a family-office context, an unsupported statement can misstate a fund term, combine reporting periods, invent a rationale or provide a confident answer where evidence is missing. The control objective is evidence-grounded output with visible limits.

Controls include approved retrieval collections, claim-level citations, answer abstention, deterministic number locks, evaluation sets, reviewer sign-off and exception reporting. The interface should make “insufficient approved evidence” an acceptable result. A higher answer rate is not automatically a better result.

9.2 Data privacy and confidentiality

Family-office data can identify family members, employees, counterparties and advisers. It can also reveal ownership structures, investments, security arrangements and personal plans. The UAE Government's official overview of Federal Decree Law No. 45 of 2021 describes processing controls, duties to secure personal data, data-subject rights and cross-border transfer requirements [14]. Applicability and any free-zone rules require legal review.

Controls begin with data classification and purpose. The office should identify what each use case needs, which data categories are excluded, where processing occurs, who can access the data and how long it is retained. Test environments should use masked or synthetic data where appropriate. Production access should follow least privilege. Public or consumer tools should not receive confidential information unless the office has explicitly approved the architecture, contract and use case.

9.3 Prompt injection and malicious content

NIST describes direct and indirect prompt injection as relevant attack modes for GenAI systems [10]. Indirect injection can enter through a document, webpage, email or other content that instructs the model to ignore system rules, reveal information or take an action.

The system should treat retrieved content as untrusted data. Instruction hierarchy, content isolation, tool permissioning, allow-listed actions, output validation and transaction separation reduce the attack surface. Sensitive tools should require structured arguments and authorisation outside the model. Security testing should include hostile documents and cross-user data-exfiltration attempts.

9.4 Stale, conflicting or incomplete data

A correct summary of an obsolete report can still mislead. Private-market valuations, ownership structures and commitments change at different frequencies. The retrieval layer should enforce effective-date logic, expose staleness and prefer approved current versions. Conflicts should remain visible until resolved.

Controls include version status, superseded-record links, maximum-age rules by field, completeness flags and source precedence. The answer should identify the oldest material data point and any missing source. Where an office uses a management estimate, the label should appear in the answer and downstream report.

9.5 Model and evaluation risk

Model behaviour can change with provider updates, prompt templates, retrieval configuration and tool definitions. A production use case needs a fixed evaluation set that includes normal, edge, incomplete and adversarial cases. Tests should cover factual support, numbers, permissions, abstention, prompt injection, sensitive-data handling and escalation.

Release management should record model and configuration versions. A material change triggers regression testing and approval. Production monitoring samples outputs and compares quality over time. A rollback path should be documented.

9.6 Third-party and concentration risk

The Financial Stability Board identifies third-party dependencies and service-provider concentration among AI-related vulnerabilities in finance [12]. Family offices can depend on cloud, model, portfolio, data and integration providers. A failure, pricing change, policy change or acquisition can affect several workflows at once.

Vendor diligence should cover financial stability, architecture, subcontractors, locations, security certifications, incident history, model training use, retention, deletion, encryption, access, service levels, change notice, audit rights, exit support and data portability. The office should know which critical workflows share the same provider and whether a manual fallback exists.

9.7 Cyber and fraud risk

The FSB identifies cyber risk and notes that GenAI can expand attack opportunities and facilitate fraud or disinformation [12]. NCSC and CISA guidance covers secure design, development, deployment, operation and maintenance [11]. The office's threat model should include credential theft, deepfake-enabled impersonation, malicious documents, data poisoning, unauthorised retrieval, tool abuse and supply-chain compromise.

Fraud controls remain independent of model confidence. A persuasive voice or message does not establish identity. High-risk requests require verification through approved channels, known contact information and segregated authorisation.

9.8 Human factors

Over-reliance can occur when users accept fluent output without checking evidence. Under-reliance can cause staff to recreate work and bypass the approved system. Training should therefore cover capability, limits, sensitive-data rules, citation checking, escalation and incident reporting.

The office should preserve challenge. Reviewers need time, access to sources and accountability for their section. If the workflow measures only speed, users may learn to approve quickly. Quality and control measures belong in performance design.

Risk	Example	Preventive control	Detective or corrective control
Confabulation	Invented fund term	Approved retrieval and abstention policy	Citation review and evaluation sampling
Privacy	Personal data sent to an unapproved service	Data classification and provider restrictions	Data-loss monitoring and incident response
Prompt injection	Document instructs the model to reveal records	Content isolation and tool allow-list	Adversarial tests and anomalous-access alerts

Risk	Example	Preventive control	Detective or corrective control
Stale data	Old private-fund value treated as current	Effective-date and version filters	Staleness flag and exception queue
Numerical error	Narrative changes an approved percentage	Locked calculation payload	Automated number comparison
Permission failure	User receives another entity's documents	Authoritative identity and pre-retrieval filters	Access-log review and canary tests
Third-party failure	Model service unavailable	Dependency map and fallback process	Service monitoring and continuity activation
Fraud	Generated instruction changes beneficiary	Transaction-plane separation and dual approval	Independent beneficiary verification
Model drift	Provider update changes answer quality	Version control and release gates	Regression tests and rollback

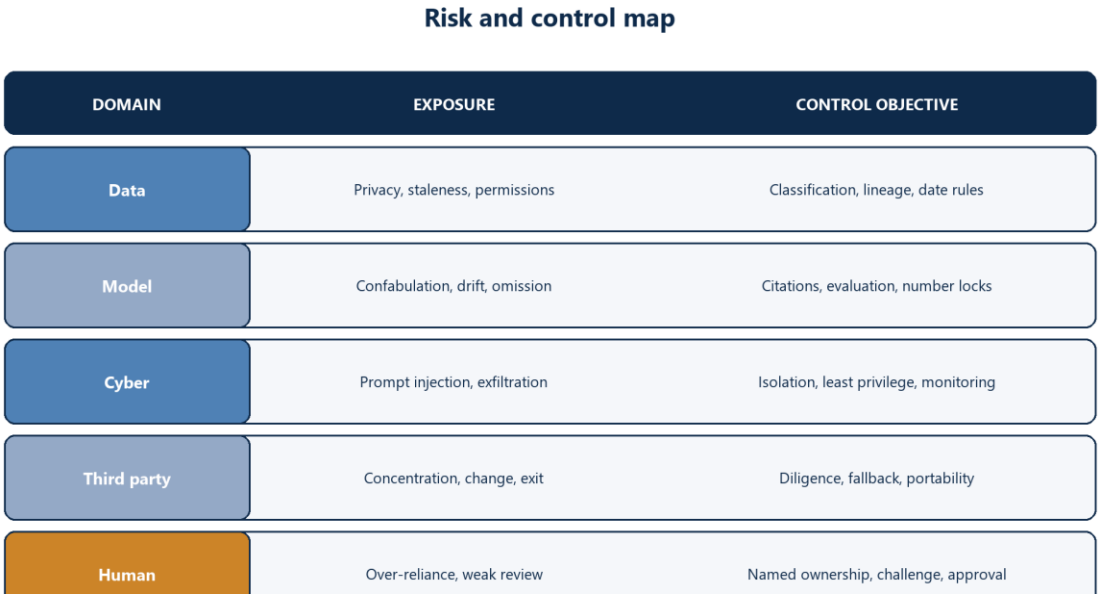

Figure 5. Risk-control map across data, model, cyber, third-party and human domains

Source: Matchpoint synthesis based on [9-16].

10. GOVERNANCE AND OPERATING MODEL

10.1 Decision rights

NIST's AI RMF organises work through Govern, Map, Measure and Manage [9]. The OECD principles emphasise privacy and human rights, transparency and explainability, robustness and security, and accountability [13]. The Matchpoint operating model translates these themes into specific decision rights.

The family principal or governing body sets risk appetite and approves material uses. An AI steering group can own the policy, inventory and prioritisation. A business owner is accountable for each use case. Data owners approve sources and access. Technology owns architecture and service operation. Security owns threat assessment and incident controls. Legal, privacy and compliance assess obligations. Independent risk or audit functions challenge the design where the office's structure supports them.

10.2 Use-case inventory

Every deployed or experimental use should be recorded. Minimum fields include purpose, owner, users, data categories, model and provider, system access, output audience, decision materiality, human review, tests, approval date, review date, incidents and retirement status.

The inventory should also capture AI embedded in existing software. Citi's report notes that family offices may access AI through established software or devices [1]. Vendor questionnaires and system reviews help identify these uses. Unrecorded embedded AI can create data and model dependencies outside the policy process.

10.3 Policy hierarchy

A concise acceptable-use policy should define approved tools, prohibited data, permitted uses, review requirements and incident reporting. Higher-materiality use cases need supporting standards for data, security, testing, vendor management, records and transactions. Procedures translate those standards into daily work.

Policy should distinguish experimentation from production. A sandbox can use masked data and no live actions. A production workflow uses approved data, identity, logging, monitoring and support. Moving between them requires evidence and approval.

10.4 Validation and assurance

Validation should be independent enough to challenge the builder. It reviews the use case, data, evaluation design, security tests, results, limits and proposed monitoring. High-materiality workflows can require external testing or specialist review.

Assurance continues after launch. Periodic review confirms purpose, users, providers, data, quality, incidents and benefit. A use case should be suspended when permissions fail, evidence quality falls below its threshold, a material provider change has not been assessed, or an incident affects its control assumptions.

10.5 Incident management

NIST's GenAI profile emphasises incident disclosure processes, and NCSC/CISA guidance includes incident management across deployment and operation [10,11]. A family-office playbook should define detection, containment, evidence preservation, assessment, notification, recovery and lessons learned.

Incident categories can include data exposure, unauthorised access, incorrect material output, prompt injection, fraudulent request, provider compromise, model failure and policy breach. Reporting routes should be simple enough for staff to use quickly. External notification requirements require legal assessment.

11. BUY, BUILD OR OUTSOURCE

11.1 Decision dimensions

Citi's 2026 report describes buy versus build as a central dilemma for family offices, reflecting constrained development resources and a fragmented provider market [1]. The decision should be made by use case and architectural layer.

Buying can accelerate access to mature workflow features and support. Building can provide tighter integration, data control and tailoring. Outsourcing can supply specialist operating capacity and governance support. Hybrid designs are common: an office can buy a portfolio platform, build a controlled retrieval layer and outsource model-security testing.

Decision criteria include strategic differentiation, data sensitivity, integration complexity, control requirements, internal skills, time to deploy, scale, provider concentration, exit cost and total lifecycle cost. A lower subscription price can be outweighed by integration, review, data remediation or migration effort.

Dimension	Buy	Build	Outsource
Time to initial capability	Often shorter for standard workflows	Longer where foundations are absent	Depends on provider onboarding
Fit to distinctive process	Configurable within product limits	Highest potential tailoring	Defined by service scope
Internal engineering need	Product ownership and integration	Product, engineering, security and support	Vendor and service governance
Data control	Contract and architecture dependent	Greater design control	Contract, personnel and environment dependent
Model flexibility	Product roadmap dependent	Higher component choice	Provider method dependent
Operational support	Vendor-supported	Office-supported or separately contracted	Included within service design
Exit complexity	Data export and replacement	Internal maintenance and technical debt	Knowledge transfer and service transition
Principal risk	Lock-in and hidden embedded AI	Capacity, security and technical debt	Dependency, confidentiality and accountability

11.2 Vendor diligence questions

The office should ask:

- Which legal entity provides each service, and which subcontractors process data?
- Where are prompts, documents, embeddings, outputs, backups and logs stored?
- Is customer content used to train or improve any shared model?
- How are tenants isolated and access privileges administered?
- Which model versions are used, and how are changes communicated?
- What evaluation and security testing is performed before release?
- How are prompt injection, data exfiltration and tool abuse addressed?
- Which records can the office export, and in which format?
- How does deletion propagate through caches, indexes, backups and logs?
- What incident notice, audit evidence and remediation rights are provided?
- Which business-continuity arrangements apply to the service?
- How can the office continue or exit if the provider changes terms or ceases service?

Answers should be evidenced through contracts, architecture documents, independent reports, tests and demonstrations. Marketing statements alone are insufficient for material workflows.

12. MATCHPOINT IMPLEMENTATION ROADMAP

12.1 Phase 0: mandate and inventory

The governing sponsor defines objectives, risk appetite and excluded uses. The team inventories current AI use, systems, providers, data stores and critical workflows. It identifies personal, confidential, investment and transaction data. It records applicable legal and contractual constraints for qualified review.

Deliverables include a use-case register, data map, provider map, initial policy and ownership model. A programme should not proceed to live sensitive data until owners and permitted uses are defined.

12.2 Phase 1: data and control foundation

The team selects one bounded document collection and one deterministic dataset. It resolves identities, permissions, source metadata, versioning and retention. Security creates the threat model. The team defines the evaluation set and acceptance thresholds before testing the model.

A suitable first use can be manager-letter search and summarisation for one asset class. It is read-only, source-verifiable and disconnected from transaction systems. The pilot collection should contain representative formats, outdated versions, conflicting statements and access boundaries.

12.3 Phase 2: controlled pilot

The pilot tests retrieval, citations, abstention, numerical fidelity, permissions, prompt injection and reviewer workflow. Baseline performance is measured using the current manual process. Production-like monitoring and incident reporting are tested even if the environment is a sandbox.

The pilot passes only when it meets pre-agreed quality and control gates. A compelling demonstration is not a substitute for a fixed evaluation. Failed cases are retained and analysed.

12.4 Phase 3: reporting integration

The system connects to approved deterministic outputs through a read-only interface. It drafts commentary using locked numbers and cited documents. Finance and portfolio reviewers approve their sections. Publication records retain the evidence package, generated draft, corrections and approvals.

The team compares total process time, reviewer time, corrections, unsupported claims and user outcomes with the baseline. Where review cost remains high, the use case is redesigned, narrowed or stopped.

12.5 Phase 4: portfolio intelligence

The office extends the canonical model across asset classes and entities. It introduces query planning, look-through status, liquidity scenarios and committee-pack generation. Each expansion repeats data, security, evaluation and approval gates.

The intelligence layer exposes evidence and limitations. High-materiality questions route to named reviewers. Transaction systems remain segregated.

12.6 Phase 5: controlled orchestration

After mature operation, the office may allow the system to open tasks, request missing documents, route exceptions or prepare structured instructions. Tools remain allow-listed and scoped. Every external communication or record change follows defined approval. Payment and trade release remain in deterministic control systems.

Gate	Required evidence	Decision
Purpose	Named owner, users, outcome and excluded uses	Approve, revise or reject use case
Data	Source map, permissions, quality, dates and retention	Admit approved collections only
Security	Threat model, access tests, adversarial tests and incident route	Proceed when residual risk is accepted
Quality	Fixed evaluation set, evidence coverage, number fidelity and abstention	Proceed only above approved thresholds
Operations	Support, monitoring, fallback, change and exit plans	Authorise controlled production
Value	Baseline and end-to-end process evidence	Scale, redesign or retire

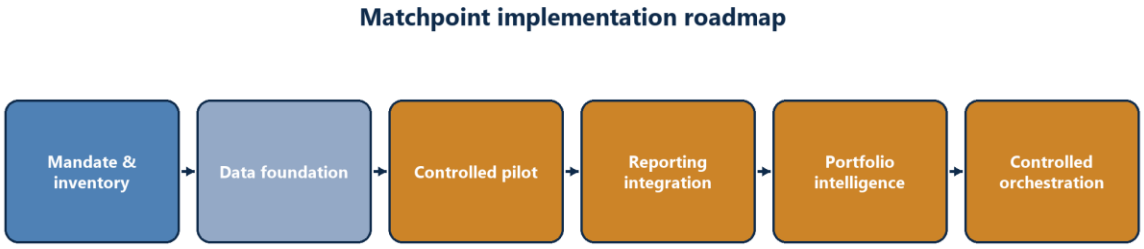

Figure 6. From mandate to controlled orchestration: phased gates and evidence packages
Source: Matchpoint scenario analysis. Timing and resources require separate office approval.

13. MATCHPOINT SCENARIO ANALYSIS

13.1 Scenario status

This section is Matchpoint scenario analysis. It is illustrative and is not observed productivity, a forecast, an investment recommendation or evidence that GenAI will improve returns. Each office should replace the assumptions with approved data.

13.2 Illustrative pilot

Consider a family office that reviews manager letters and quarterly reports across a defined private-markets portfolio. The current process requires analysts to locate documents, identify portfolio-company developments, compare prior-period statements and prepare a committee summary.

The illustrative pilot covers one quarter, one approved document collection and a read-only interface. It excludes email, personal data, payment systems and external communications. The system retrieves cited passages, identifies changes and drafts a structured summary. Analysts verify every material statement and record corrections.

The baseline measures documents processed, elapsed time, reviewer time, missed issues, rework and evidence completeness. The pilot records the same measures plus unsupported-answer rate, citation correctness, permission failures, abstention and security-test results. Value is recognised only when the complete controlled process improves against the baseline and meets its quality gates.

13.3 Scale decision

The office scales the pilot if evidence coverage, accuracy, permissions, review effort and user outcomes meet approved thresholds. It redesigns the pilot if retrieval or data quality causes recurring exceptions. It retires the use case if control or review costs exceed the demonstrated operating value.

No financial-return benefit is assumed. The scenario concerns information-processing quality, cycle time and control evidence.

14. LIMITATIONS AND RESEARCH AGENDA

The family-office evidence base remains limited. Survey samples, definitions and fieldwork periods differ. Self-reported AI use does not reveal workflow coverage, architecture, control quality or realised benefit. Interviews provide detailed observations and do not establish prevalence.

The technology and provider landscape changes rapidly. Model behaviour, service terms, data practices and regulations require current review at implementation. The NIST AI RMF 1.0 is under revision at the source-review date [9]. Local legal applicability and regulatory perimeter require specialist advice.

Future research should examine independently measured family-office productivity, error rates, reviewer burden, privacy incidents, vendor concentration and the long-term effect of AI on staffing and governance. Comparable evaluation datasets for investment reporting and private-markets documents would improve evidence quality. Research should also distinguish retrieval quality, calculation quality, narrative quality and decision outcomes.

15. CONCLUSION

Generative AI is becoming relevant to family-office operations through reporting, document intelligence and information retrieval. Survey and interview evidence shows growing adoption and persistent barriers in expertise, privacy, security and value measurement [1-4]. Broader financial-sector guidance identifies data quality, model risk, cyber risk, third-party concentration and accountability as central concerns [6,9-17].

The operating design determines whether these capabilities become reliable decision support. A governed data foundation establishes sources, dates, ownership, lineage and permissions. Deterministic systems retain calculation, valuation and transaction authority. Controlled retrieval supplies approved evidence. The generative layer prepares cited explanations and drafts. Named people validate material outputs and authorise actions.

The practical programme begins with a bounded, read-only and verifiable use case. It establishes the baseline, evaluation set, security tests, monitoring and incident process before production. Expansion proceeds through explicit gates. This route turns GenAI from an informal drafting tool into an auditable information capability suited to family capital.

DECLARATIONS

Evidence boundary. Survey findings describe the cited samples and fieldwork periods. They do not establish causation, implementation quality or investment performance.

Scenario disclosure. Matchpoint scenario analysis is illustrative. It is not observed productivity, a forecast, an investment recommendation or evidence that GenAI will improve returns.

Professional-advice boundary. This paper is general research. It is not legal, tax, accounting, cybersecurity or investment advice. Readers should obtain qualified advice for their circumstances and jurisdictions.

Conflicts and funding. Matchpoint Partners prepared this paper as independent research for its MP Insights programme. No external research funding or vendor consideration is represented in the evidence register.

Source-review date. Sources and linked materials were reviewed on 31 July 2026.

JEL Classification: G11, G23, G32, M15, O33 **Keywords:** generative AI, family office, portfolio intelligence, investment reporting, private markets, retrieval-augmented generation, data governance, AI risk management, human oversight

REFERENCES

[1] Citi Institute and Citi Wealth (2026). *AI in the Family Office: Privacy, Efficiency and Institutional Rigor*. May 2026. https://on.citi/ai_in_the_family_office_pdf

[2] Citi Wealth (2025). *2025 Global Family Office Report*. https://www.privatebank.citibank.com/doc/family-office/2025-global-family-office-report.pdf

[3] Deloitte Private (2024). *Digital Transformation of Family Office Operations, 2024*. The Family Office Insights Series, Global Edition. https://www.deloitte.com/global/en/services/deloitte-private/research/family-office-insights-series.html

[4] UBS (2025). *Global Family Office Report 2025*. https://advisors.ubs.com/mediahandler/media/742678/GFOreport.pdf

[5] J.P. Morgan Private Bank (2026). *2026 Global Family Office Report*. https://privatebank.jpmorgan.com/nam/en/insights/reports/2026-family-office-report

[6] Dubai Financial Services Authority (2025). *New DFSA AI survey: Generative AI adoption has nearly tripled within the DIFC in last 12 months as governance continues to develop*. 12 November 2025. https://www.dfsa.ae/news/new-dfsa-ai-survey-generative-ai-adoption-has-nearly-tripled-within-difc-last-12-months-governance-continues-develop

[7] Institutional Limited Partners Association (2025). *ILPA Reporting Template (v. 2.0) Suggested Guidance*. January 2025. https://ilpa.org/wp-content/uploads/2025/01/ILPA-Reporting-Template-v.-2.0-Suggested-Guidance.pdf

[8] IFRS Foundation. *IFRS 13 Fair Value Measurement*. https://www.ifrs.org/issued-standards/list-of-standards/ifrs-13-fair-value-measurement/

[9] Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. https://doi.org/10.6028/NIST.AI.100-1

[10] National Institute of Standards and Technology (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1. https://doi.org/10.6028/NIST.AI.600-1

[11] UK National Cyber Security Centre, US Cybersecurity and Infrastructure Security Agency, and international partners (2023). *Guidelines for Secure AI System Development*. https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf

[12] Financial Stability Board (2024). *The Financial Stability Implications of Artificial Intelligence*. 14 November 2024. https://www.fsb.org/uploads/P14112024.pdf

[13] OECD.AI (2024). *OECD AI Principles overview*. Updated May 2024. https://oecd.ai/en/principles

[14] Official Portal of the UAE Government (2024). *Data protection laws*. Updated 22 October 2024. https://u.ae/en/about-the-uae/digital-uae/data/data-protection-laws

[15] International Monetary Fund (2024). *Global Financial Stability Report*, Chapter 3, "Advances in Artificial Intelligence: Implications for Capital Market Activities". October 2024. https://www.imf.org/-/media/Files/Publications/GFSR/2024/October/English/textrevised.ashx

[16] Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V. and Stein, M. (2024). *Intelligent financial system: how AI is transforming finance*. BIS Working Papers No 1194. https://www.bis.org/publ/work1194.htm

[17] International Organization of Securities Commissions (2025). *Artificial Intelligence in Capital Markets: Use Cases, Risks and Challenges*. IOSCOPD788. https://www.iosco.org/library/pubdocs/pdf/IOSCOPD788.pdf

APPENDIX A. USE-CASE CONTROL RECORD

A.1 Minimum fields

Field	Purpose
Use-case name and ID	Stable inventory reference
Business owner	Accountable person for outcome and review
Purpose and users	Approved scope
Prohibited uses	Explicit boundary
Data categories and sources	Privacy, confidentiality and lineage
Provider, model and version	Dependency and change control
System permissions	Authority boundary
Output audience	Publication and confidentiality control
Human reviewers	Named validation roles
Evaluation set and thresholds	Release and monitoring evidence
Security tests	Threat-model evidence
Retention and deletion	Records and privacy control
Approval and review dates	Governance evidence
Incidents and exceptions	Ongoing risk record
Retirement plan	Exit and data-disposition control

A.2 Output labels

Outputs should carry one or more labels:

- **Verified source fact:** directly supported by an approved source.
- **Deterministic calculation:** produced by an approved engine with an identifier.
- **Manager-reported:** attributed to an external manager and not independently verified by the office.
- **Management estimate:** provided or approved by management; methodology and date stated.
- **Generated draft:** requires review before use.
- **Insufficient evidence:** no decision-grade answer available from approved sources.

APPENDIX B. PILOT EVALUATION SCORECARD

Dimension	Example measure	Required interpretation
Evidence coverage	Supported material claims divided by total material claims	Higher is better; define materiality in advance
Citation correctness	Citations that directly support the associated statement	Check source and passage
Numerical fidelity	Numbers matching approved payload or source	Any material mismatch requires correction
Omission	Required items absent from the output	Evaluate against a reference checklist

Dimension	Example measure	Required interpretation
Abstention	Unsupported questions declined or escalated	Measure appropriate abstention, not answer volume
Permission control	Cross-user and cross-entity access tests passed	Any failure is a security event
Adversarial resilience	Prompt-injection and malicious-document cases handled	Retain failed cases for regression testing
Reviewer effort	Time and corrections required for approval	Compare end to end with baseline
Operational reliability	Availability, latency and retrieval failure	Compare with process service level
Incident readiness	Detection, escalation and recovery exercise	Evidence includes owners and elapsed response

APPENDIX C. VENDOR DILIGENCE RECORD

The vendor file should retain the proposal, architecture, data-flow map, security evidence, privacy terms, subprocessors, model-change process, evaluation evidence, service levels, incident commitments, exit terms and approved exceptions. Material verbal statements should be confirmed in writing. Contract review requires qualified legal advice.

APPENDIX D. GLOSSARY

Canonical data model: A controlled representation that maps different source systems into consistent entities, assets, fields and definitions.

Confabulation: Plausible generated content that is false, unsupported or inconsistent with available evidence; NIST uses this term in its GenAI risk taxonomy [10].

Deterministic analytics: Calculations that follow defined code, rules and inputs and can be reproduced independently.

Embedding: A numerical representation used to compare semantic similarity for retrieval and other machine-learning tasks.

Generative AI: A class of AI systems that creates content such as text, code, images or audio in response to inputs.

Human-in-the-loop: A workflow in which a person reviews, approves or provides feedback at defined decision points.

Prompt injection: An attack or manipulation in which direct user input or retrieved content attempts to override instructions or induce unsafe behaviour.

Retrieval-augmented generation: A workflow that retrieves selected information and supplies it to a generative model for an answer or draft.

System of record: The authoritative source for a controlled record or state.

System of intelligence: A layer that combines approved records, analytics and context to support questions and decisions.