International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

LLM-Generated Executive Briefings: Automating 10-K & Earnings-Call Analysis for Enterprise Sellers

Adish Rai

Account Manager
Amazon Web Services Inc., USA
rai.adish@gmail.com

Abstract:
Outbound noise is at an all-time high and generic outreach underperforms. Recent benchmarks put cold-email replies and cold-call success in the low single digits, pushing enterprise sellers toward deeper, company-specific context. This paper proposes a practical, single-pass long-context approach that ingests a target company's latest 10-K/10-Q and earnings-call transcript, then produces a seller-ready brief and 2 - 3 persona-specific draft emails. Each recommendation includes a verbatim quote with a precise source reference (timestamp or filing section) to ground the demand hypothesis. We detail an AWS-aligned implementation (Bedrock + Lambda + S3) and a lightweight evaluation rubric (factuality, utility, and operational metrics). We also discuss a data-fusion extension that incorporates CRM and conversation-intelligence data (e.g., meeting notes) to tailor outputs to an account's live opportunities. A brief sidebar outlines optional enhancements (ephemeral chunking and retrieval) for teams that need lower latency/cost at scale. While the reference design is AWS-centric, the pattern is portable to any stack with equivalent services.

Keywords: Sales enablement; earnings calls; 10-K; large language models; Amazon Bedrock; CRM; enterprise sales

1. INTRODUCTION

Cold outbound is increasingly inefficient: multiple 2024 - 2025 benchmarks cite low single-digit reply or success rates for cold email and cold calling [1], [2]. The implication is straightforward: sellers who demonstrate mastery of a prospect's strategic initiatives, risks, and priorities stand out.

1.1 The Enterprise Seller's Constraint

Enterprise AEs/AMs typically own large books of business - dozens to hundreds of named accounts across segments and regions. A given fiscal quarter can produce hundreds of pages of new information per account: a 10-K (~40k words on average), interim 10-Qs, investor decks, and an earnings call (~30 - 60 minutes) [3] - [6]. In parallel, sellers must manage pipeline hygiene, internal forecasting, renewal motions, and live customer meetings. In practice, only a small subset of accounts receives deep research; the remainder receives generic outreach, which underperforms.

1.2 Signal Buried in Financial Narratives

Filings and earnings calls surface strategic themes (e.g., platform modernization, margin protection), explicit risks (e.g., concentration, regulatory), initiatives (e.g., data-platform migration), and spend signals (e.g., capex guidance). These are the raw materials for demand hypotheses - clear, testable statements about how a seller's solution can address a current priority. However, translating long, finance-oriented narratives into persona-specific, solution-aligned messages remains slow and error-prone.

1.3 Paper Contribution

We describe a minimal, production-credible pipeline that converts 10-K/10-Q filings and an earnings-call transcript into a cited seller brief and persona emails in one long-context LLM pass - no heavy retrieval infrastructure required. We then extend this design with an account-context fusion approach that adds CRM fields (industry, stage, open opportunities) and conversation-intelligence highlights (e.g., recent meeting notes) to further tailor outputs. We provide an AWS-aligned implementation pattern, evaluation metrics, and governance considerations.

2. RELATED WORK AND BACKGROUND

2.1 Data Access

The SEC provides programmatic access to filings and XBRL via EDGAR; teams should observe fair-use guidance and rate limits. Earnings-call transcripts are not hosted by EDGAR and are typically accessed through commercial APIs (e.g., Finnhub, Financial Modeling Prep), which return cleaned text and sometimes speaker/timestamp metadata [3] - [5].

2.2 Model Context

Models available via Amazon Bedrock (e.g., Anthropic Claude family) support ~200k-token contexts - roughly 150k words, 500+ pages - sufficient for a single 10-K plus the latest call transcript in many cases [7]. Some newer models provide extended contexts up to ~1M tokens; these can accommodate multiple documents but may introduce latency and cost trade-offs relative to a single long-context pass [8], [9].

2.3 Analyst Platforms

Research tools provide high-quality generic earnings summaries but generally do not tailor outputs to a vendor's specific product plays or a particular account's CRM context - gaps addressed here using a simple domain prompt, mandatory citations, and an optional account-context fusion step.

3. PROBLEM STATEMENT (SALES REALITY AND REQUIREMENTS)

We formalize the challenge along four dimensions that a practical system must address:

3.1 Coverage Under Time Pressure

AEs and AMs cannot allocate an hour per account to listen to calls and parse filings during earnings season. A credible system must deliver useful briefs within minutes, not hours, and should minimize human review effort by attaching exact citations.

3.2 Relevance to Solution and Persona

Generic summaries are not enough. Outputs must map company themes to vendor capabilities to buyer personas (e.g., CIO cares about platform risk/ROI; VP Data about performance/SLA).

3.3 Trust and Auditability

Inference must be grounded. Every recommendation should include a verbatim quote and a precise locator (timestamp or Item/page), enabling instant verification and reducing hallucination risk.

3.4 Extensibility to Account Context

Insights should integrate with CRM and meeting intelligence to reflect what has already been discussed, current stage, and next steps. This prevents tone-deaf emails, avoids duplication, and supports multi-threading across a buying committee.

4. REFERENCE ARCHITECTURE (SIMPLE LONG-CONTEXT DESIGN)

Objective: Produce a seller brief and persona emails from a company's latest filings and earnings call - each recommendation backed by a quote and source reference.

4.1 Inputs

- Company capabilities (short catalog): pains solved, outcomes, proof points, target personas.
- Sources: (a) 10-K/10-Q (text/HTML via EDGAR), (b) earnings-call transcript (text via a transcript API) [3] - [5].

4.2 Single LLM Pass (with cite-or-skip policy)

Provide the capability catalog, personas, and the full transcript (optionally the 10-K/10-Q) to a long-context model (e.g., Claude via Bedrock), with the following rule:

All claims must include a verbatim quote (≤25 words) and a source_ref: for calls, a timestamp (e.g., 00:21:13 - 00:21:40); for filings, an Item + page/section (e.g., "10-K Item 7, p. 42"). If a claim lacks evidence, omit it.

Model outputs (structured):

- Brief: strategic_themes[], risks[], initiatives[] - each with why_it_matters_to_us, evidence_quote, source_ref.
- Emails: 2 - 3 short drafts per persona; each includes evidence_quote + source_ref.

4.3 Delivery and Human-in-the-Loop

Deliver outputs to Slack and/or attach to the account/opportunity in CRM. Reps approve/edit before sending. Attach or link the full transcript (and optionally the 10-K) for rapid cross-check.

5. IMPLEMENTATION PATTERN (AWS-ALIGNED)

5.1 Data Fetch and Storage

A scheduled AWS Lambda (via Amazon EventBridge) calls the EDGAR API for the latest 10-K/10-Q and a transcript provider for the most recent call; raw sources are stored in Amazon S3 (encrypted with KMS) [3].

5.2 Generation

A Lambda function calls Amazon Bedrock (e.g., Claude 3) with: (a) capability catalog, (b) personas, (c) full transcript, (d) optional 10-K text, and (e) a strict JSON schema with a cite-or-skip guardrail. CloudWatch Logs capture prompts/outputs for audit [7].

5.3 Delivery

A Slack webhook (from Lambda) posts the seller brief and persona emails; a Salesforce REST call can attach the JSON and a formatted note to the relevant account/opportunity.

5.4 Licensing

Verify transcript-provider redistribution terms before forwarding full text outside the system of record.

5.5 Portability

The pattern is cloud-agnostic. Any platform with equivalent primitives (object storage, scheduled jobs, long-context LLM, and messaging/CRM APIs) can implement the same approach.

6. ACCOUNT-CONTEXT FUSION (CRM AND CONVERSATION INTELLIGENCE)

6.1 Motivation

The same strategic theme can imply different outreach depending on account status. If Salesforce indicates an active opportunity with Security as a stakeholder, the system should draft an email for the CISO rather than the CIO, referencing the earnings quote and prior call notes.

6.2 Minimal Data Fields

The fusion step reads only lightweight, non-PII fields, such as: (1) industry/sub-industry; (2) opportunity stage and primary product; (3) last-meeting highlights (1 - 3 bullets) from conversation intelligence; (4) ICP tags (e.g., data-platform focus, cloud provider). These fields condition generation without exposing sensitive data.

6.3 Mapping Logic

Maintain a small capability catalog (YAML/JSON). For each capability, list typical pains, outcomes, personas, and proof points. Maintain a short list of trigger keywords that are likely to appear in filings or calls (for example: "margin," "modernization," "SLA," "migration," "security posture"). The model (or a simple rules pass) matches the extracted themes to capabilities, then selects a persona template based on account context.

6.4 Governance
All CRM and notes fields are encrypted at rest, passed minimally to the model, and never echoed beyond what is necessary for the draft. Logging is configured to mask PII at capture (for example, names and email addresses are replaced with [REDACTED]) before any record is written to storage. Admins can disable fusion on sensitive accounts.

7. OUTPUT FORMAT
To keep review simple and consistent, outputs follow a short, fixed structure:

Seller brief (1 - 1.5 pages):
- 3 - 5 strategic themes.
- For each theme: one-sentence explanation of why it matters, plus a short evidence quote and a precise source reference (earnings-call timestamp or filing Item/page).
- Optional: 1 - 3 risks and 1 - 3 initiatives, each with an evidence quote and source reference.

Persona emails (2 - 3 drafts):
- Each draft is 90 - 130 words, tailored to a specific persona (for example, CIO, VP Data, CISO).
- Each includes the evidence quote and the source reference to justify the message.

8. EVALUATION FRAMEWORK
Factual accuracy. (a) Citation coverage - the share of themes/emails that include both an evidence quote and a source reference; (b) Human check pass rate - reviewers confirm that the cited text truly supports the claim.

Usefulness for sellers. (a) Time saved - average minutes of research saved per account; (b) Usefulness rating - reps rate each brief/email from 1 to 5; (c) Action rate - the share of drafts that reps edit/approve and send; (d) (optional) Outcome lift - changes in replies or meetings booked, noting that many factors beyond this system affect outcomes.

Operations (speed and cost). (a) Latency per brief - measure the median (P50) and the 95th percentile (P95); (b) Token cost per run - with and without the 10-K included; (c) Throughput during peak earnings weeks.

Suggested test design. Split accounts into a treatment group (receives the brief/emails) and a control group (business-as-usual). Compare the groups on the metrics above, focusing on medians to be robust to outliers, and run for 4 to 6 weeks.

9. RISKS, PRIVACY, AND GOVERNANCE
- Licensing and redistribution. Transcript vendors vary in terms; confirm what you may attach or forward.
- Over-confident generation. The cite-or-skip rule blocks unsupported claims; logs enable audit.
- PII and CRM data. Pass only necessary fields to the model; encrypt at rest (S3/KMS); redact logs; restrict IAM roles.
- Model drift. Version prompts and the capability catalog; re-validate outputs periodically and after model upgrades.

10. ENHANCEMENTS FOR SCALE AND RELIABILITY (OPTIONAL SIDEBAR)
A single long-context pass is often enough. When documents get very long or you need faster, cheaper runs, two simple add-ons help:

1) Ephemeral chunking (split, then select). Break the 10-K into natural sections (Items 1, 1A, 7, 7A). Break the call transcript by section/speaker. Quickly select the few sections that look relevant (by keywords or a

quick semantic search). Send only those sections to the model. This reduces tokens and speeds things up, and it still allows precise citations because section boundaries are preserved.

2) Temporary retrieval ("RAG-lite"). Compute temporary embeddings for the sections (for example, using Titan Embeddings on Bedrock). Pick the top 10 - 20 most relevant sections compared to your capability catalog and the personas you are targeting. Use them as context for generation, then delete the temporary data. No persistent vector database is required.

11. CONCLUSION

In a market where undifferentiated outbound underperforms, sellers win by demonstrating specific, evidenced understanding of a prospect's priorities. A simple, long-context pipeline - EDGAR filings + the latest earnings-call transcript to one guardrailed LLM pass to a cited brief and persona-specific draft emails - delivers immediate, repeatable value without heavy ML infrastructure. The account-context fusion extension further tailors outreach using lightweight CRM and meeting-intelligence fields. Teams can later layer in ephemeral chunking/retrieval for cost and latency gains.

REFERENCES:

[1] Belkins, "What are B2B Cold Email Response Rates? (2025 Study)," Jul. 9, 2025. [Online]. Available: https://belkins.io/blog/cold-email-response-rates. Accessed: Aug. 5, 2025.
[2] Cognism, "The Top Cold Calling Success Rates for 2025 Explained," Mar. 18, 2025. [Online]. Available: https://www.cognism.com/blog/cold-calling-success-rates. Accessed: Aug. 5, 2025.
[3] U.S. Securities and Exchange Commission, "EDGAR Application Programming Interfaces (APIs)," Jun. 6, 2024. [Online]. Available: https://www.sec.gov/search-filings/edgar-application-programming-interfaces. Accessed: Aug. 5, 2025.
[4] Finnhub, "Earnings Call Transcripts API," 2025. [Online]. Available: https://finnhub.io/docs/api/earnings-call-transcripts-api. Accessed: Aug. 5, 2025.
[5] Financial Modeling Prep, "Earnings Transcript API," 2025. [Online]. Available: https://site.financialmodelingprep.com/developer/docs/earning-call-transcript-api. Accessed: Aug. 5, 2025.
[6] Encyclopædia Britannica, "Earnings Conference Calls," 2024. [Online]. Available: https://www.britannica.com/money/earnings-conference-calls. Accessed: Aug. 5, 2025.
[7] Amazon Web Services, "Anthropic's Claude in Amazon Bedrock," 2024. [Online]. Available: https://aws.amazon.com/bedrock/anthropic/. Accessed: Aug. 5, 2025.
[8] Google, "Introducing Gemini 1.5, Google's next-generation AI model," Feb. 15, 2024. [Online]. Available: https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/. Accessed: Aug. 5, 2025.
[9] Google, "Long context | Gemini API," 2025. [Online]. Available: https://ai.google.dev/gemini-api/docs/long-context. Accessed: Aug. 5, 2025.