



Innovate.DTCC: AI-Powered Hackathon

Self-Service Investigation (Chat NLP)
Corporate Action Reference Data Ingestion

PRESENTER: Team InfyTelligence

DATE: February 7, 2025

COMPANY: Infosys

Agenda



- 1. Team Introduction**
- 2. Problem Statement**
- 3. Solution Overview**
- 4. Technical Implementation**
- 5. Model Performance and Metrics**
- 6. Industry Impact & Collaboration**
- 7. Demo**
- 8. Conclusion and Q&A**
- 9. Appendix**

Team Introduction

Use Case Title: AI-Powered Corporate Action Data Ingestion and Processing

Team InfyTelligence



Nikhil Goel
Corporate Action SME



Zeynep Ozdas
Corporate Action SME



Jasleen Singh
Specialist AI Engineer



Soumya Chakraborty
Senior Technical Architect



Ranjith Kumar Baskaran
Specialist AI Engineer



Appasaheb Sirsat
Technical Manager



Sharmeen Javed Shaikh
Senior Generative AI Developer



Anuja Datta
Technology Lead



Yashvant Sridharan
Generative AI Developer



Anil Kumar Nagabhiaru
Senior AI Architect

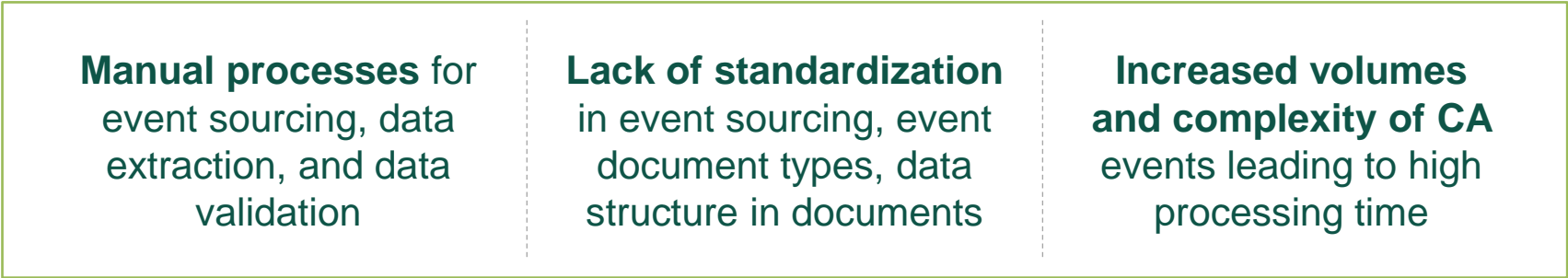
Problem Statement

The Challenge

Corporate Actions Volume



Complexity around capturing of CA events information



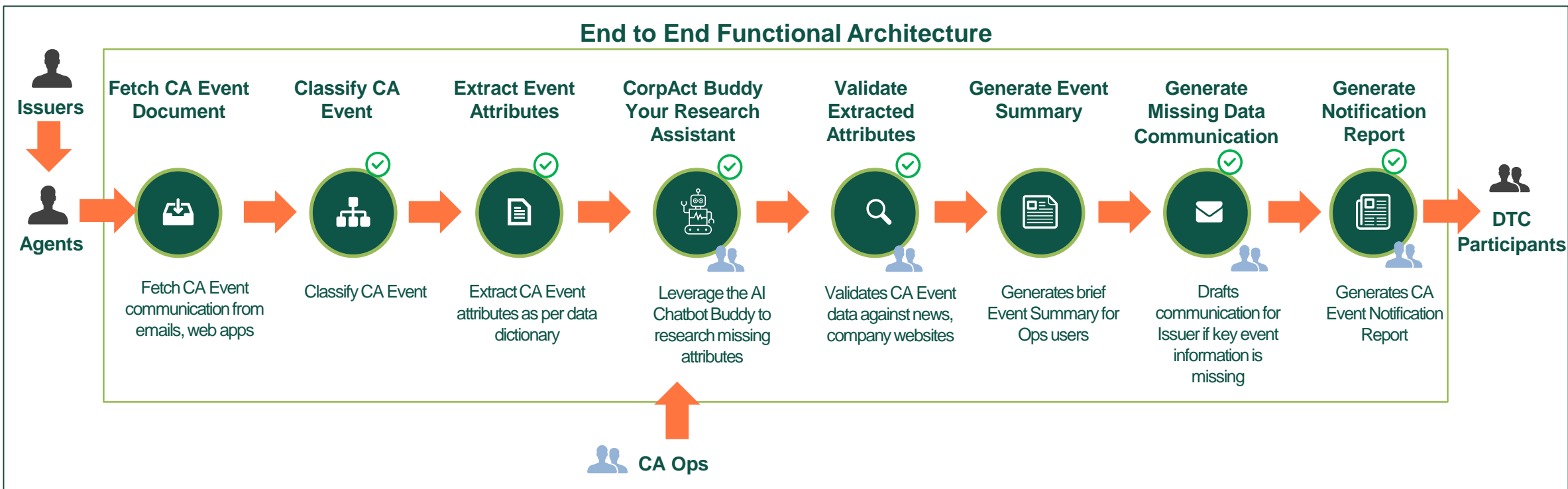
Industry Need

The industry must **automate** corporate action event sourcing to streamline operations, minimize risk, and ensure timely access to accurate data

Solution Overview

Our Approach

A scalable Generative AI-based system for ingesting, extracting, and validating corporate actions data



Key Features

End to End
CA Event Sourcing
Life Cycle

GenAI
Solution

Operational
Efficiency with
Increased Accuracy

In-scope CA Events

Full
Call

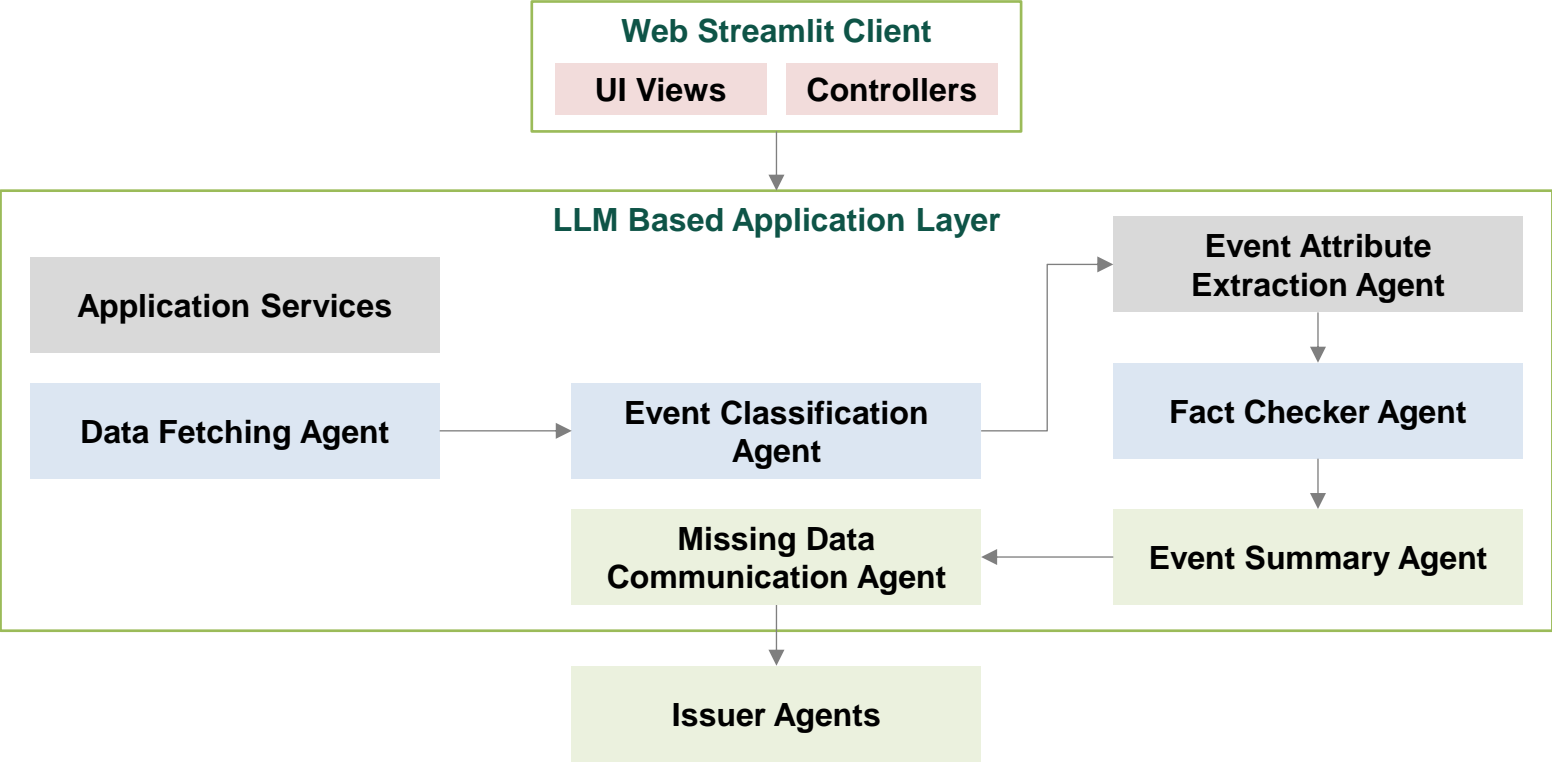
Partial
Call

Merger

✔ Highlighted items are in-scope for Hackathon

Technical Implementation

Technical Architecture – AI solution to Ingest Corporate Action Reference Data



Tech Stack

- Programming: Python
- Storage: AWS S3
- Deployment: AWS EC2
- LLMs: claude-3-5-sonnet

Data Usage

- We used publicly available dataset (SEC regulatory filings)
- Leveraged Python libraries to pre-process regulatory filings
- Defined corporate actions contextual prompts to extract key event attributes using LLM calls

Applicable for Hackathon Implementation

Scalability	Accuracy	Maintenance	Implementation Complexity	Software Cost	Operational Costs	Responsible AI
High	Very High	Low	Low	Low	Low	N/A

Model Performance and Metrics

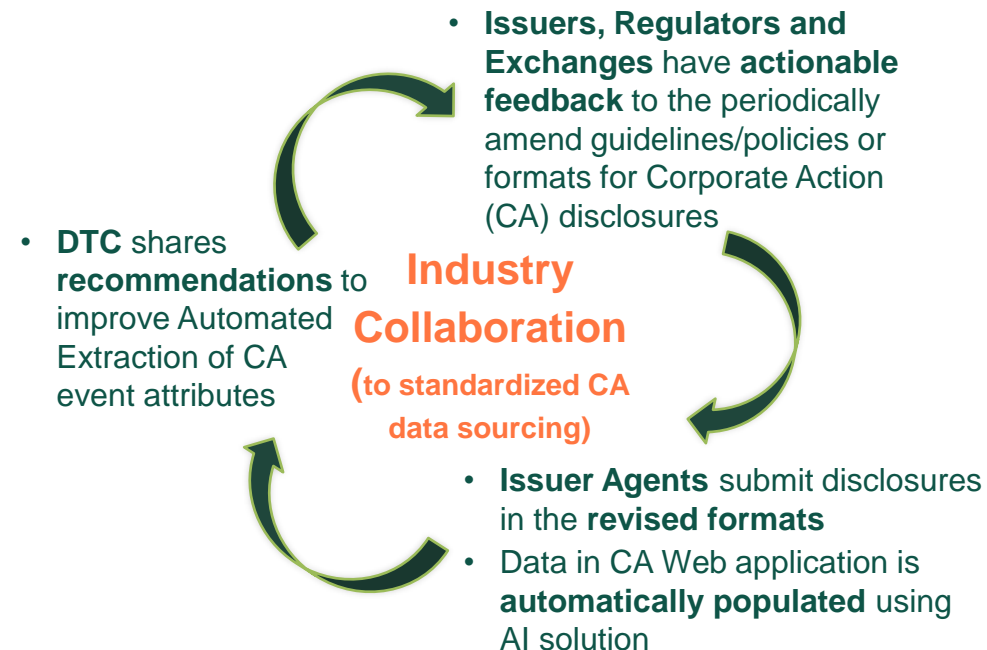
Metric Category	Metric Name	Performance Value	Comments
Accuracy & Data Quality	Precision	97%	<ul style="list-style-type: none">Observed for in-scope CA events with limited data
	Recall	100%	<ul style="list-style-type: none">Observed for in-scope CA events with limited data
	F1-Score	99%	<ul style="list-style-type: none">Observed for in-scope CA events with limited data
	False Positive Rate (FPR)	11%	<ul style="list-style-type: none">Room to improve prompts for extracted attributes/false positives
Efficiency & Automation Metrics	Processing Time per Filing	70% to 80% reduction in processing time	<ul style="list-style-type: none">Observed for in-scope CA events with limited data
Cost & Resource Optimization Metrics	Reduction in Manual Processing Costs	60% to 70% reduction in manual processing cost*	<ul style="list-style-type: none">Observed for in-scope CA events with limited data

* This includes manual processing cost, bottleneck (maker-checker) costs, penalty due to operational errors costs

Industry Impact, Collaboration and Production Viability

Path to Production

Measure	Hackathon Solution	Production Solution	Path to Production - Key Improvement Areas
Scalability	Medium	High	<ul style="list-style-type: none"> Leverage Cloud for Scalability and API for integrations
Accuracy	High	Very High	<ul style="list-style-type: none"> Implement Autonomous Agents Introduce Multi-model Capabilities
Maintenance	Low	Medium	<ul style="list-style-type: none"> LLM + RAG requires minimal fine-tuning
Implementation Complexity	Low	Medium	<ul style="list-style-type: none"> Leverage Agentic RAG to optimize LLM usage Improve UX/UI for intuitive user experience
Software Cost	Low	Medium	<ul style="list-style-type: none"> Leverage Agentic RAG to optimize LLM usage
Operation Cost	Low	Low	<ul style="list-style-type: none"> Implement automated prompt to research missing attributes using CorpAct Buddy connected LLM have access to real-time data
Responsible AI	NA	Medium	<ul style="list-style-type: none"> Implement controls around sending sensitive data to LLM/ CorpAct Buddy Implement Fact Check Agent for to avoid bias and data accuracy
Data Source	SEC 8K	Issuer Notice, Exchange Filing, News	<ul style="list-style-type: none"> Expand scope to source/validate data using Issuer Notices, Exchange Filings, News etc. SEC EDGAR Integration



Challenges and Limitations

- Documents containing multiple CA events, event updates
- Complex documents containing tables/images
- Industry resistance to standardized document/data language
- CorpAct Buddy is not connected to LLM; accessing to real-time data

Solution Demo



Demo Steps

-  **Upload CA Event Regulatory Filing**
Upload 8K SEC Filing
-  **Event Classification**
Classify CA Event
-  **Event Data Extraction**
Extract CA Event attributes as per data dictionary
-  **Human-in-the-loop and CorpAct Buddy**
CA Ops review/ edit/ update event attributes, research using CorpAct Buddy
-  **Event Summary Generation**
Generates brief Event Summary for Ops users
-  **Missing Data Communication**
Drafts communication for Issuer Agent if key Event information is missing
-  **Notification Report Generation**
Generates CA Event Notification Report

Conclusion

Key Metrics

Business Impact

- We observed **70 to 80 %** faster processing than traditional methods

Financial Efficiency

- This will help to cut corporate action processing costs by **60% to 70%**

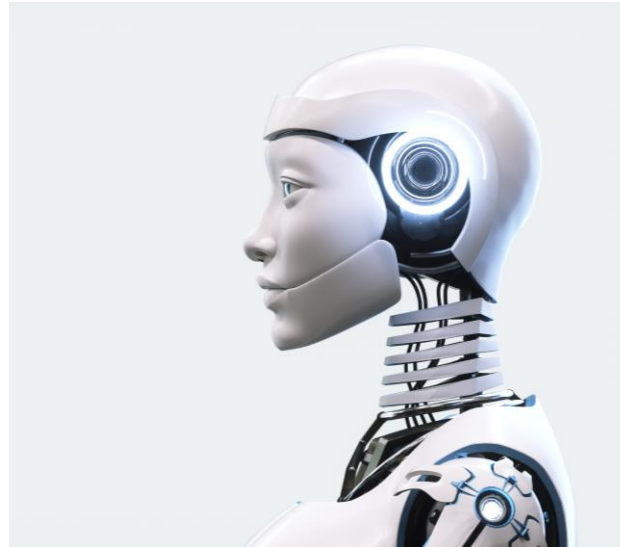
Accuracy & Data Quality

- We observed high degree of accuracy and data quality – Precision – **97 %**, Recall – **100 %** and F1-Score – **99 %**



AI Opportunity

Key metrics confirm that **Gen AI Solution** holds **tremendous potential to automate** the extraction of corporate actions information from documents.



Industry Benefits



Promotes Straight Through Processing



Reduction in Operations Cost



Reduction in Operational Risk



Enhanced Accuracy and Timelines



Industry Collaborations

Thank You

Financial Markets.

Forward.TM

Appendix



1. Metrics Measurement
2. Application Screenshots

Raw Measurement

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	CA Event - Full Call								CA Event - Partial Call									CA Event - Merger								
2	Full Call - Marina Joint Powers Financing	Attributes from Tool	M	Data from Tool	Available	Test	Result		Partial Call - American Pacific Corporation	Attributes from Tool	M	Data from Tool	Available	Test	Result	Remarks		Merger - Cash/Securities_SEC_AT&T	Attributes from Tool	M	Data from Tool	Available	Test	Result		
3	BaseCusip	basecusip	M	568087		Y	Y	True Positive	BaseCusip	basecusip	M	28740	Y	Y	True Positive			CAEvent	caevent	M	Merger	Y	Y	True Positive		
4	Class	class	M	568087AM2		Y	Y	True Positive	Class	class	M	028740 AF5	Y	Y	True Positive			CASubEvent	casubevent	M	Cash and Securities	Y	Y	True Positive		
5	ConditionalPaymentApplicableFlag	conditionalpaymentapplicableflag	M	No		Y	Y	True Positive	ConditionalPaymentApplicableFlag	conditionalpaymentapplicableflag	M	No	Y	Y	True Positive			AcquiringCompany	acquiringcompany	M	AT&T Inc.	Y	Y	True Positive		
6	ContactE-mail	contactemail	M	Not Available		Y	Y	True Negative	ContactE-mail	contactemail	M	Not Available	Y	Y	True Negative			TargetCompany	targetcompany	M	Time Warner Inc.	Y	Y	True Positive		
7	ContactPhoneNumber	contactphonenumber	M	Not Available		Y	Y	True Negative	ContactPhoneNumber	contactphonenumber	M	1-800-344-5128	Y	Y	True Positive			AnnouncementDate	announcementdate	M	22-Oct-16	Y	Y	True Positive		
8	Currency	currency	M	USD		Y	Y	True Positive	Currency	currency	M	USD	Y	Y	True Positive			RecordDate	recorddate	M	Not Available	Y	Y	True Negative		
9	CUSIP	cusip	M	568087AM2		Y	Y	True Positive	CUSIP	cusip	M	028740 AF5	Y	Y	True Positive			EffectiveDate	effectivedate	M	Not Available	Y	Y	True Negative		
10	CAEvent	caevent	M	Full Call		Y	Y	True Positive	CAEvent	caevent	M	Partial Call	Y	Y	True Positive			PaymentDate	paymentdate	M	Not Available	Y	Y	True Negative		
11	CAEventCategory	caeventcategory	M	Redemptions		Y	Y	True Positive	CAEventCategory	caeventcategory	M	Redemptions	Y	Y	True Positive			ExchangeRatio	exchangeratio	M	Parent Stock	Y	Y	True Positive		
12	IssuerName(s)	issuername	M	Marina Joint Powers Fin		Y	Y	True Positive	IssuerName(s)	issuername	M	American Pacific Corporation	Y	Y	True Positive			CashAmount	cashamount	M	\$53.75 per share	Y	Y	True Positive		
13	SecuritySymbol	securitysymbol	M	Not Available		Y	Y	True Negative	SecuritySymbol	securitysymbol	M	Not Available	Y	Y	True Negative			DealValue	dealvalue	M	Not Available	Y	Y	True Negative		
14	Maturity	maturity	M	11/15/2036		Y	Y	True Positive	Maturity	maturity	M	2015	Y	Y	True Positive			Additions / Premiums	additionspremiums	M	Not Available	Y	Y	True Negative		
15	OutstandingNumberOfSecurities	outstandingnumberofsecurities	M	Not Available		Y	Y	True Negative	OutstandingNumberOfSecurities	outstandingnumberofsecurities	M	10,000,000	Y	N	False Positive	this is \$ value but ok		TargetCompanyOwnershipDistributionPercentage	targetcompanyownershippercentage	M	Not Available	Y	N	False Negative		
16	Premium/ CashRate	premiumcashrate	M	No premium		Y	Y	True Positive	Premium/ CashRate	premiumcashrate	M	2.25%	Y	Y	True Positive			CombinedPrimaryExchange	combinedprimaryexchange	M	NYSE	Y	Y	True Positive		
17	RedemptionPrice	price	M	\$15,000.00		Y	N	False Positive	RedemptionPrice	price	M	102.25%	Y	Y	True Positive			VotingRequired	votingrequired	M	Yes	Y	Y	True Positive		
18	PublicationDate / DatedDate / RecordDate	publicationdatedatedaterecorddate	M	4/11/2016		Y	Y	True Positive	PublicationDate / DatedDate / RecordDate	publicationdatedatedaterecorddate	M	9-Aug-12	Y	Y	True Positive			Currency	currency	M	USD	Y	Y	True Positive		
19	Rate	rate	M	3.95%		Y	Y	True Positive	Rate	rate	M	9%	Y	Y	True Positive			CUSIP/ ISIN/ RIC/ SEDOL	cusipisinricsedol	M	Not Available	Y	Y	True Negative		
20	RedemptionAmount	redemptionamount	M	\$15,000.00		Y	Y	True Positive	RedemptionAmount	redemptionamount	M	40,000,000	Y	Y	True Positive	\$ sign should be there										
21	RedemptionDate	redemptiondate	M	5/15/2016		Y	Y	True Positive	RedemptionDate	redemptiondate	M	8/9/2012	Y	Y	True Positive											
22	SubIssueType	subissuetype	M	Municipal Bonds		Y	Y	True Positive	SubIssueType	subissuetype	M	Corporate Bond	Y	Y	True Positive											
23	TrusteeAgent/PayingAgent	trusteeagentpayingagent	M	MUFG Union Bank, N.A.		Y	Y	True Positive	TrusteeAgent/PayingAgent	trusteeagentpayingagent	M	Wells Fargo Bank, National	Y	Y	True Positive											
24																										
25	Full Call - Enovix Corp	Attributes from Tool	M	Data from Tool	Available	Test	Result		Partial Call - Coca Cola	Attributes from Tool	M	Data from Tool	Available	Test	Result	Remarks		Merger - Stock_SEC_AMR	Attributes from Tool	M	Data from Tool	Available	Test	Result		
26	BaseCusip	basecusip	M	293594		Y	Y	True Positive	BaseCusip	basecusip	M	Not Available	Y	Y	True Negative			CAEvent	caevent	M	Merger	Y	Y	True Positive		
27	Class	class	M	293594115		Y	Y	True Positive	Class	class	M	2.375% Senior Notes due 2018	Y	Y	True Positive			CASubEvent	casubevent	M	Stock-for-stock	Y	Y	True Positive		
28	ConditionalPaymentApplicableFlag	conditionalpaymentapplicableflag	M	No		Y	Y	True Positive	ConditionalPaymentApplicableFlag	conditionalpaymentapplicableflag	M	No	Y	Y	True Positive			AcquiringCompany	acquiringcompany	M	US Airways Group	Y	Y	True Positive		
29	ContactE-mail	contactemail	M	Not Available		Y	Y	True Negative	ContactE-mail	contactemail	M	manadyla.castro@kof.com.mx, jesus.calleja@kof.com.mx	Y	Y	True Positive			TargetCompany	targetcompany	M	AMR Corporation	Y	Y	True Positive		
30	ContactPhoneNumber	contactphonenumber	M	(800) 736-3001, +1(781) 57		Y	Y	True Positive	ContactPhoneNumber	contactphonenumber	M	(5255) 1519-5186, (5255) 1519-5216, (5255) 1519-5013	Y	Y	True Positive			AnnouncementDate	announcementdate	M	14-Feb-13	Y	Y	True Positive		
31	Currency	currency	M	USD		Y	Y	True Positive	Currency	currency	M	USD	Y	Y	True Positive			RecordDate	recorddate	M	Not Available	Y	Y	True Negative		
32	CUSIP	cusip	M	293594115		Y	Y	True Positive	CUSIP	cusip	M	Not Available	Y	Y	True Negative			EffectiveDate	effectivedate	M	Expected in third q	Y	Y	True Positive		
33	CAEvent	caevent	M	Full Call		Y	Y	True Positive	CAEvent	caevent	M	Partial Call	Y	Y	True Positive			PaymentDate	paymentdate	M	Not Available	Y	Y	True Negative		
34	CAEventCategory	caeventcategory	M	Redemptions		Y	Y	True Positive	CAEventCategory	caeventcategory	M	Redemptions	Y	Y	True Positive			ExchangeRatio	exchangeratio	M	one share of comm	Y	Y	True Positive		
35	IssuerName(s)	issuername	M	Enovix Corporation		Y	Y	True Positive	IssuerName(s)	issuername	M	Coca-Cola	Y	Y	True Positive			CashAmount	cashamount	M	Not Available	Y	Y	True Negative		
36	SecuritySymbol	securitysymbol	M	ENOVIX		Y	Y	True Positive	SecuritySymbol	securitysymbol	M	KOFL, KOF	Y	Y	True Positive			DealValue	dealvalue	M	\$11 billion	Y	N	False Positive		
37	Maturity	maturity	M	Not Available		Y	Y	True Negative	Maturity	maturity	M	2018	Y	Y	True Positive			Additions / Premiums	additionspremiums	M	Not Available	Y	Y	True Negative		
38	OutstandingNumberOfSecurities	outstandingnumberofsecurities	M	Not Available		Y	Y	True Negative	OutstandingNumberOfSecurities	outstandingnumberofsecurities	M	US\$1,000,000,000.00	Y	N	False Positive	this is \$ value but ok		TargetCompanyOwnershipDistributionPercentage	targetcompanyownershippercentage	M	72%	Y	Y	True Positive		
39	Premium/ CashRate	premiumcashrate	M	\$0.01 per Public Warrant		Y	Y	True Positive	Premium/ CashRate	premiumcashrate	M	Not Available	Y	Y	True Negative			CombinedPrimaryExchange	combinedprimaryexchange	M	Not Available	Y	Y	True Negative		
40	RedemptionPrice	price	M	\$11.50 per share		Y	Y	True Positive	RedemptionPrice	price	M	Not Available	Y	Y	True Negative			VotingRequired	votingrequired	M	US Airways shares	Y	N	False Positive		
41	PublicationDate / DatedDate / RecordDate	publicationdatedatedaterecorddate	M	7-Dec-21		Y	Y	True Positive	PublicationDate / DatedDate / RecordDate	publicationdatedatedaterecorddate	M	12-Jul-17	Y	Y	True Positive			Currency	currency	M	USD	Y	Y	True Positive		
42	Rate	rate	M	Not Available		N	Y	True Negative	Rate	rate	M	2.375%	Y	Y	True Positive			CUSIP/ ISIN/ RIC/ SEDOL	cusipisinricsedol	M	Not Available	Y	Y	True Negative		
43	RedemptionAmount	redemptionamount	M	Not Available		N	Y	True Negative	RedemptionAmount	redemptionamount	M	US\$55,000,000	Y	Y	True Positive											
44	RedemptionDate	redemptiondate	M	7/7/2022		Y	Y	True Positive	RedemptionDate	redemptiondate	M	08/18/2017	Y	Y	True Positive											
45	SubIssueType	subissuetype	M	Corporate Bond		Y	Y	True Positive	SubIssueType	subissuetype	M	Corporate Bond	Y	Y	True Positive											
46	TrusteeAgent/PayingAgent	trusteeagentpayingagent	M	CompuShare Trust Co		Y	Y	True Positive	TrusteeAgent/PayingAgent	trusteeagentpayingagent	M	The Bank of New York Mellon	Y	Y	True Positive											

Overall Scores

CA Metrics																	
Events	# of documents tested	# of attributes in event	# of attributes extracted	# of attributes tested	% Completeness	# of accurate attributes	% Accuracy	Time Taken to Record (in Min)	True Positive- TP	False Positive - FP	True Negative - TN	False Negative- FN	Precision = TP/(TP+FP)	Recall = TP/(TP+FN)	F1 Score = 2×Precision×Recall / (Precision+Recall)	FPR (False Postive Ratio) = FP / (FP+TN)	
Full Call	4	84	84	84	100%	82	98%	99	66	2	16	0	97%	100%	99%	11%	
Doc 1 - Enovix Corp	1	21	21	21	100%	20	95%	24	16	0	5	0	100%	100%	100%	0%	
Doc 2 - Marina Joint Powers Financing Authority	1	21	21	21	100%	20	95%	28	16	1	4	0	94%	100%	97%	20%	
Doc 3 - Qytera Technologies	1	21	21	21	100%	21	100%	25	17	0	4	0	100%	100%	100%	0%	
Doc 4 - CITY OF ARLINGTON, TEXAS	1	21	21	21	100%	21	100%	22	17	1	3	0	94%	100%	97%	25%	
Partial Call	3	63	63	63	100%	59	94%	80	51	3	8	0	94%	100%	97%	27%	
Doc 1- American Pacific Corporation	1	21	21	21	100%	20	95%	30	18	1	2	0	95%	100%	97%	33%	
Doc 2- Coca Cola	1	21	21	21	100%	20	95%	30	16	0	4	0	100%	100%	100%	0%	
Doc 3- Hertz	1	21	21	21	100%	19	90%	20	17	2	2	0	89%	100%	94%	50%	
Merger	4	68	68	68	100%	63	93%	102	36	2	27	3	95%	92%	94%	7%	
Doc 1- Merger_Cash&Securities_SEC_AT&T	1	17	17	17	100%	16	94%	30	10	0	6	1	100%	91%	95%	0%	
Doc 2 - Merger_Stock_SEC_AMR	1	17	17	17	100%	15	88%	26	9	2	6	0	82%	100%	90%	25%	
Doc 3 - Merger_Stock_SEC_Disney	1	17	17	17	100%	16	94%	24	8	0	8	1	100%	89%	94%	0%	
Doc 4 - Merger_Cash_SEC_MSFT	1	17	17	17	100%	16	94%	22	9	0	7	1	100%	90%	95%	0%	
Overall Score	11	215	215	215	100%	204	95%	281	153	7	51	3	96%	98%	97%	12%	
Industry Benchmark																	
F1 Score																	
Good F1 Score: In most real-world applications, an F1 score around 0.85 - 0.95 is considered excellent for automated extraction of structured corporate action data from well-structured documents (e.g., ISO 20022 messages, spreadsheets, etc.).																	
Moderate F1 Score: For less structured or noisy data (e.g., press releases, unstructured text), a 0.70 - 0.85 F1 score may be achievable, but performance will likely vary depending on how clean and standardized the data is.																	
Lower F1 Score: For complex, highly ambiguous documents with non-standardized formats, F1 scores below 0.70 may be expected, but these can be improved with additional fine-tuning and better preprocessing techniques.																	
FPR Score																	
Excellent Performance (Low FPR): An FPR of less than 5% is generally considered excellent in financial document extraction. A system with such a low FPR is good at correctly identifying corporate action events without flagging too many non-events.																	
Good Performance (Moderate FPR): A moderate FPR range of 5% - 10% is often seen in many practical implementations of corporate action extraction systems. This range still allows for high precision and avoids overwhelming users with too many false positives.																	
Acceptable Performance (High FPR): An FPR above 10%, especially in systems designed for high-volume processing, may indicate that the model or extraction process needs tuning or improvements, as a higher FPR could significantly increase the number of false alarms.																	

Application Screenshots (1/3)

Innovate DTCC: AI-Powered Hackathon

AI-Powered Corporate Action Data Ingestion

1. Upload CA event Documents

Choose one ZIP file or multiple PDF files

Drag and drop files here
Limit 200MB per file • ZIP, PDF

Browse files

PartialCall_CorporateBond_Nasdaq_Bombardier.pdf

Processed 1 files successfully!

CA Event Documents

Document ID	Trigger	File Name	CA Event	Issuer Name	Uploaded Date	Uploaded By
101	Issuer Document	PartialCall_CorporateBond_Nasdaq_Bombardier.pdf	Partial Call	Bombardier Inc	02/06/2025	User

2. Extract Document

Enter Document ID to Extract:

101

Extract

Application Screenshots (2/3)

3. Partial Call Processing

06/02/2025, 23:01

kof20170712_6k.htm - Generated by SEC Publisher for SEC Filing

6-K 1 kof20170712_6k.htm KOF20170712_6K

UNITED STATES

SECURITIES AND EXCHANGE COMMISSION

Washington, D.C. 20549

FORM 6-K

REPORT OF FOREIGN PRIVATE ISSUER

PURSUANT TO RULE 13a-16 OR 15d-16 UNDER

THE SECURITIES EXCHANGE ACT OF 1934

For the month of July 2017

Commission File Number 1-12260

COCA-COLA FEMSA, S.A.B. de C.V.

(Translation of registrant's name into English)

United Mexican States

(Jurisdiction of incorporation or organization)

	Attribute Name	Attributes Type	Extracted Value
1	basecusip	M	Not Available
2	class	M	2.375% Senior Notes due 2018
3	conditionalpaymentapplic	M	No
4	contactemail	M	mariadyla.castro@kof.com.mx,
5	contactphonenumber	M	(5255) 1519-5186, (5255) 1519-5
6	currency	M	USD
7	cusip	M	Not Available
8	caevent	M	Partial Call
9	caeventcategory	M	Redemptions
10	issuename	M	Coca-Cola FEMSA, S.A.B. de C.V.
11	securitysymbol	M	KOFL, KOF
12	maturity	M	2018
13	outstandingnumberofsecu	M	555,000,000

Chat with CorpAct Buddy

Welcome back, How can I assist you with your Corporate Actions Research today?

Application Screenshots (3/3)

4. Missing Data Communication Email

Issue summary:

Issuer: Bombardier Inc.

Security Name: Corporate Bond

Event Type: Call

Upon review, we have identified the following mandatory data points that are currently missing:

- cusip
- outstandingnumberofsecurities
- premiumcashrate

To ensure accurate processing and timely communication to stakeholders, please provide the missing details at your earliest convenience.
Let us know if further clarification is needed.

Copy