keystone.ai | Novartis

MMF AI Proposal

Status: 11/26 Working Draft

Contents

Engagement Overview

The Commercial AI Opportunity & Challenge

Novartis's commercial organization wants to transform how it leverages data to power best-in-class AI models to improve efficiency, inform better decision-making, and drive meaningful increases in operating profitability.

There's an opportunity to leverage a new class of breakthrough "AI" models that can be deployed for relevant use cases for Novartis. Examples include generalized deep-learning models for sequence-to-sequence prediction that can be leveraged for an array of forecasting tasks (GTN, long-range, production planning), LLM based processing models that can extract terms from unstructured text to configure logic-based automation workflows, causal inference for quantifying uplift of marketing pend or contract-impact analysis.

These models work best on using the finest-grain data possible (e.g., claim-level dispense data, specific terms from EDIs, etc.,) and preparing data in the right format, particularly in commercial pharmaceutical settings posed an array of challenges.

Keystone believes it has the capabilities, specifically the RAIN Data Platform and experience to develop a platform that addresses these core challenges, solving which, has the ability to unlock value and opportunity across the entire commercial and operations organizations at Novartis.

Keystone proposes an engagement with Novartis to deploy several MVP models for 3 Managed Markets Finance use cases and in the process deploy its RAIN data platform. In doing so, Keystone will deploy several key capabilities reusable across use cases and allow Novartis to evaluate Keystone capabilities for longer-term transformation.

<u>A starting point: 3 Managed Market Finance "MMF" Use Cases</u>

Through preliminary scoping discussions, Novartis has detailed 3 initial use cases where we can deploy RAIN data platform capabilities combined with several discrete models.

1. **Chargeback Process Automation:** Use process automation, LLM based-term extraction, and external / internal API calls to resolve "Customer not found" chargeback kickouts from Model N.
2. **Medicaid GTN Forecasting:** Deploy Keystone Topline forecasting algorithms to improve accuracy of Medicaid gross-to-net forecasts, specifically by modelling relationship between claim level-dispense data and sellout data (867).
3. **Medicaid Rebate Automation:** Analyze past Novartis Medicaid invoices and claim-level utilization data and deploy a claim processing model to triage and accelerate invoice review, particularly the review of sub-500K claims.

This proposal contains scope detail for a 16-week initial engagement with explicitly defined scope / goals / milestones for each of the above use cases. (The scope for phase 1A is explicitly defined in Exhibit A)

Summary of Scope & Timeline

The table below summarizes key deliverables / milestones for a 16-week engagement. Each milestone deliverable is described in detail in the "Sequencing & Milestones" section of each scope deep-dive. (Sections 4, 5, and 6 of this document)

The overarching goal of phase 1A, is to deploy MVP workflow automation and models on Novartis environment that prove value and meet stated success criteria "phase 1A" goals defined in each scope section. This will involve ingesting, processing, and encoding data in Keystone's RAIN data platform. Phase 1B would focus on "productionizing" those models, deploying integrations between operational systems, and building on those MVP models additional improvements.

Note on timeline: We'd expect the timeline for each individual workstream to be approximately 3 – 3.5 months. Running each workstream simultaneously in parallel, regardless of resourcing, will incur some cross-workstream dependencies. To account for these, we've built in an additional 2 – 4 weeks into the timeline to extend the timeline for phase 1B to 4 months.

Scope	Milestone Deliverable	Deliverable Type	Target Date
PHASE 1A			**~16 Weeks[1]**
1 **Chargeback Process Automation**	[1.1] Prototype Automation Workflow	Executable Source Code | Keystone Environment	Weeks 1 – 6
	[1.2] Development / Testing Workflow	Executable Source Code | Novartis Environment	Weeks 6 – 14
	[1.3] MVP Automation Workflow	Executable Source Code | Novartis Environment	Week 14
2 **Medicaid GTN Forecasting**	[2.1] Forecasting Data & Modelling Insight	Analysis (PPT/Word)	Week 1 – 6
	[2.2] MVP Medicaid Utilization Forecasting Model	Executable Source Code | Novartis Environment	Week 12
	[2.3] Back Test of Forecasting Model	Data (Excel)	Week 14

[1] Project start (i.e., week 1) will depend on minimum system and data access requirements, which are listed in the Appendix section X.

	[2.4] GTN Impact Analysis	Analysis (PPT/Word)	Week 16
3 **Medicaid Rebate Automation**	[3.1] Claim Automation Data & Modelling Insight	Analysis (PPT/Word)	Week 1 – 6
	[3.2] LLM Term Extraction Workflow	Executable Source Code | Novartis Environment	Week 10
	[3.3] MVP Rebate Invoice Triaging Model	Executable Source Code | Novartis Environment	Week 14
PHASE 1B[2]			**8-12 Weeks**
1 **Chargeback Process Automation**	1. Proactive customer creation logic 2. Expanded automation for additional kickout processing (e.g., product not found, price discrepancies) 3. Improve edge-cases / operability / UX from phase 1 MVP.		
2 **Medicaid GTN Forecasting**	1. Productuctize existing model in reusable workflow for Novartis Deploy code as dockized executable containers in model pipeline. Implement Model orchestration and CI/CD in Git repositories. 2. Add price estimation to model framework. 3. Analyze additional casual variables.		
3 **Medicaid Rebate Automation**	1. Productuctize existing model in reusable workflow for Novartis. Deploy code as dockized executable containers in model pipeline. Implement Model orchestration and CI/CD in Git repositories. 2. Deploy automatic integrations to Model N. 3. Expand model capabilities to do claim-level-data verification, such as 340B Duplicate Check, channel trend identification. 4. Test and expand model coverage to above 500K invoices.		

While the end-product (models) for each scope are relatively discrete models, they will each draw from a common set of data platform capabilities. These data platform capabilities will be deployed through Keystone's RAIN data platform. **The high-level solution architecture is detailed in Exhibit B under RAIN Data Platform Overview.**

Keystone Team | Deployment Model | Price

[WIP] Deliberately not drafted until we have alignment on scope.

[2] The exact scope would be determined following phase 1A.

Exhibit: Phase 1A Scope Detail

Scope 1: Chargeback Process Automation

Scope Overview

The Novartis MMF ICCR team wants to reduce the manual effort to process (approve / reject) chargeback invoices from wholesalers. The immediate goal is to automate highly manual workflows, specifically the workflow to create new customers following an EDI submission. However, the broader goal is to drive E2E transformation and enhancement of chargeback processing workflows, shifting from reactive to proactive processes and significantly reducing manual effort required for highly repetitive, time-consuming, and automatable tasks.

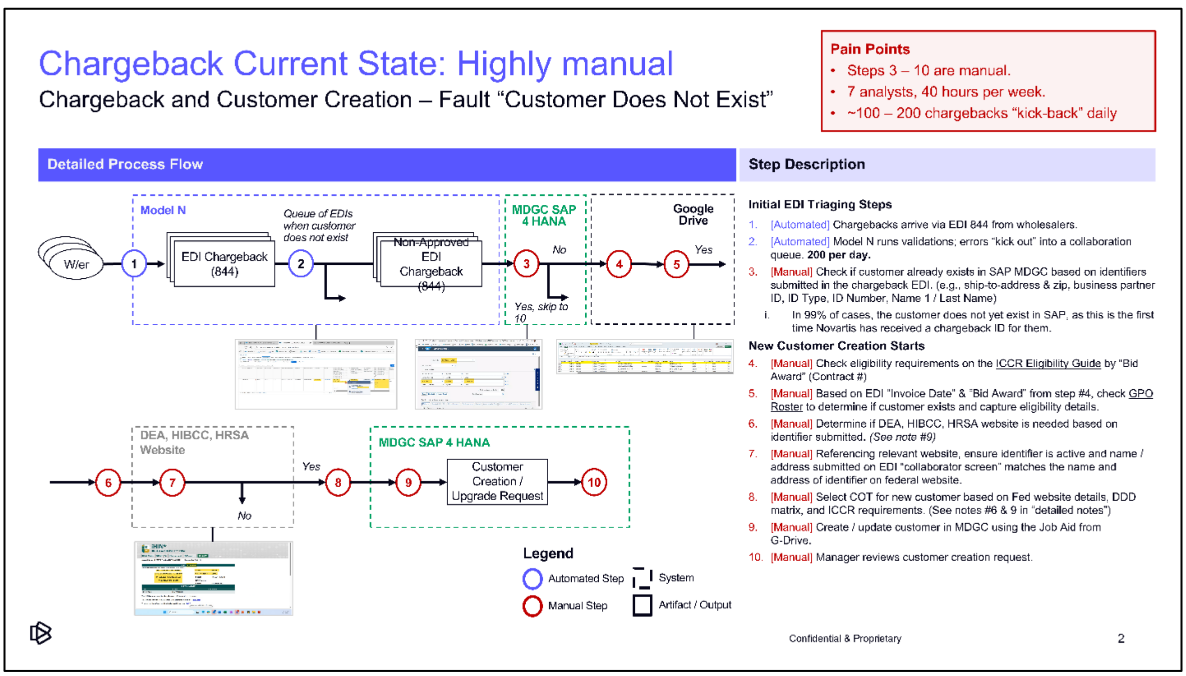
Keystone proposes an initial scope to deploy RAIN data platform and build an automated workflow to resolve “Customer not found” chargeback kickouts from Model N. In doing so, Keystone will deploy several key capabilities reusable across use cases and allow Novartis to evaluate Keystone capabilities for longer-term transformation.

Phase 1A | Solution Detail

Problem

Between 50 – 200 chargebacks are “kicked-out” of Model N due to a no customer found error each day. These chargebacks must be resolved within 5 days, but the process to process each kick-out involves 8 highly manual and repetitive steps that are currently all processed manually by 5 full-time analysts, which on average take ~25 minutes to complete. These steps covered in this process are summarized in the table below. Additional detail is in the Appendix “Additional Process Detail”.

Figure 2: Chargeback Processing Diagram – Process Steps & Systems



Solution Scope

In phase 1 of this engagement lasting ~16 weeks, Keystone will deploy its RAIN data platform on Novartis chosen cloud environment, ingest / programmatically query underlying data sources, and develop the following workflow automations. At a high-level the workflow will intake the following inputs, reference the following sources, and output the following.

- Input(s): Chargeback EDI Kicked Out due to No Customer Found Error.
- Reference Sources: (1) G-Drive Knowledge Base (including GPO Roster, DDD Matrix, ICCR Eligibility Guide), (2) Federal Website.
- Output(s): Detail for Customer Creation Request (i.e., COT, ID Number, Name, Address)

We will deploy a UX layer to allow Novartis ICCR users to initiate the workflow, review the workflow outputs, and triage EDIs that are not addressed by the workflow. All workflow artifacts (inputs, outputs, intermediary outputs, compliance screenshots, etc.,) will be accessible and indexed for easy search and retrieval. The following table outlines the specific automations (and their priority) to be deployed.

Process Step *Corresponding to Figure 2*	**Automation**	**Importance**
3	**Programmatically query SAP 4 HANA** to check if customer exists	Med
4 & 5	**Programmatically query G-Drive "knowledge base" to capture eligibility details** Ingest customer's contract eligibility specifications from chargeback EDI in Model N, query G-Drive reference docs, and retrieve customer's eligibility details (identifier ID, name, address).	Highest
6 & 7	**Programmatically determine and query appropriate federal websites to confirm contract eligibility** Ingest contract eligibility details (identifier ID, name, address) sourced in steps 4 & 5, lookup corresponding customer details from applicable federal website (DEA, HIBCC, HRSA depending on identifier), confirm details match and determine discrepancies.	Highest
8	**Automatically make class-of-trade "COT" determination** Programmatically reference G-Drive reference docs (specifically DDD matrix) and results from Federal website look-ups to make an accurate "class of trade" for each new customer.	High
9	**Programmatically create customer in SAP 4 HANA** Using the customer's eligibility details, federal website details, and class-of-trade determination, programmatically create a new customer entity in SAP 4 HANA.	Med

Phase 1A | Goals

1. **Implement a reusable workflow that drives meaningful reduction in time to process no customer kickouts**

Our primary objective is to provide "meaningful" time savings for Vanessa Zanni's ICCR team. During the first 0 – 4 weeks of prototyping we will work with Vanessa's team to align on a technically feasible and specific quantification of "meaningful", and her team will make the final determination of success.

But to establish an initial baseline, we estimate that based on the resourcing used to address no-customer-found kickouts (~5 FT analyst) and kickout volume (~500 per week), this would suggest it takes roughly 25 human minutes today to process a kickout. Based on prior experience deploying automation workflows, we would target at least at 10+ minute reduction, although again, this will be determined following prototyping.[3] [4]

It is assumed that any manual time <u>added</u> by automation – such as triaging automation outputs for accuracy or copying outputs into SAP 4 HANA – will be included when evaluating time-savings of the future state process.

2. **<u>Demonstrate reusable platform components, extensible to additional chargeback processing automation</u>**

Keystone will demonstrate to Novartis how Configure reusable components and architecture capable of scaling to additional chargeback processing task. (e.g., other Model N kickout types)

Phase 1A | Sequencing & Milestones

Project Stage	Development Milestones	Milestone Details	Target Date
Phase 1A	**[1.1] Prototype Workflow | Keystone Environment**	**What:** Experimental workflow environment on KS environment using provided data samples. **Purpose:** Environment for KS to experiment with automations 1 – 3. Demo to Novartis for initial concept feedback and use to evaluate scope of technical feasibility.	Weeks 0 – 4
	[1.2] Development Workflow | Novartis Environment	**What:** Workflow with initial versions of automations 1 – 3 deployed on Novartis environment. **Purpose:** Rapidly test, iterate, troubleshoot, and receive quick feedback on usability from Novartis team. Test specific handling with select EDIs in controlled setting with Novartis team. Target UAT for weeks 12-14.	Weeks 6 – 14
	[1.3] MVP Workflow | Novartis Environment	**What:** Automation workflow deployed on Novartis environment that meets success criteria.	Week 14
Phase 1B	**Novartis Production Workflow**	Exact scope TBD. Would include following HL items. 1. Proactive customer creation logic 2. Expanded automation for additional kickout processing (e.g., product not found, price discrepancies) 3. Improve edge-cases / operability / UX from phase 1 MVP. 4. *(Pending phase 1)* Complete interoperability with Model N and SAP 4 HANA.	Weeks 16 – 28

[3] We understand there's 5-6 ICCR analysts working full-time to resolve no customer-found kickouts, of which on average ~500 are received every week. Working 40 hours a week, this implies that each kickout takes roughly ~24 human minutes to resolve.

[4] Another benchmark metric is the full-time analysts needed to maintain the kickout queue. Currently we understand the target age between receipt and resolution of a kickout is 5 days and 6 analysts are required to maintain that age.

Scope 2: Medicaid Rebate Forecasting & Invoice Automation

Scope Overview

The MMF Medicaid rebate team wants to improve how it leverages available data to produce timely, easily reproduceable insights that can inform and improve key decision-making processes. Keystone proposes a first phase of work to incorporate claim-level-dispense data and sellout data (867) into a unified forecasting framework to help improve the accuracy of Medicaid gross-to-net forecasts.

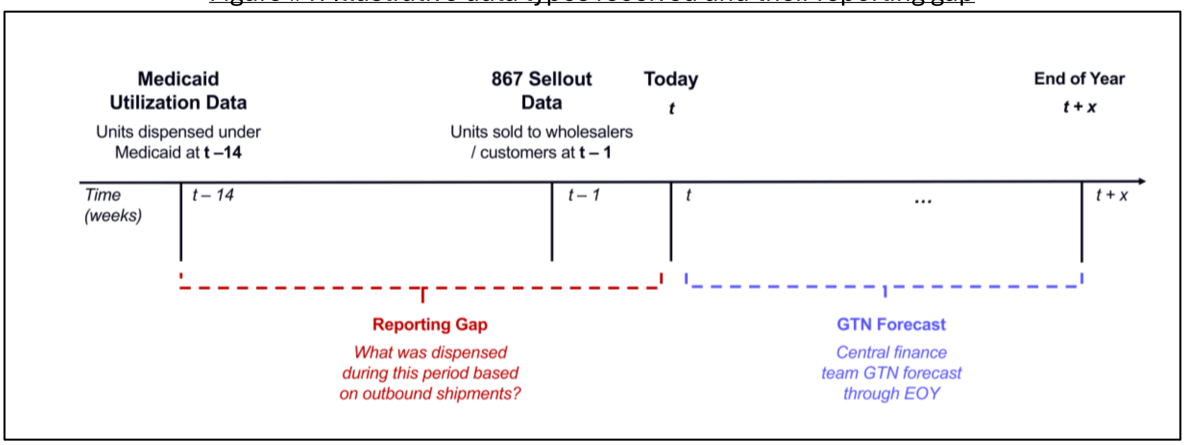
During this phase Keystone will deploy its RAIN data platform to Novartis environment, ingest key data sources, and develop a Medicaid utilization forecasting model.

Phase 1A | Solution Detail

Problem

Claim level dispensing data is received from State Medicaid entities at a significant data lag between the receipt of the data and the reported dispensing / billing date. This lag is often over 3 months, which creates a reporting gap in Medicaid utilization. This gap reduces the ability to create a gross-to-net forecast of expected Medicaid dispense volume and rebates, as at the time of forecast, we are missing the 3 or more recent months of data that can reflect key trends in utilization. This "reporting gap" is illustrated in the diagram below.

Figure #1: Illustrative data types received and their reporting gap



As an example, when the central finance team is preparing data for the July Q2 quarterly report, the central finance needs to forecast Medicaid rebates incurred through end-of-year, but utilization data is only partially available through end-of-March or Q1.

Solution

In the first phase of work, Keystone proposes addressing this reporting lag by deploying our foundation forecasting model to create a short-term Medicaid utilization "forecast" by brand.[5] Critically, this forecast will utilize leading indicators with shorter reporting lags, such as 867 sellout data, to inform estimates of Medicaid dispensing volume.

We intend to model the lead / lag relationship between 867 sellout data and Medicaid utilization data to address this reporting Gap. We will analyze this lead-lag relationships at the most granular level when possible, such as at individual provider IDs. When analyzing this relationship at the granular level, we realize there might be significant mismatches between the provider IDs present in 867 data and the

[5] In other words, this is an estimation of actuals rather than a forecast. But for simplicities sake, we will refer to this as a forecast going forward.

Medicaid utilization data. We will evaluate the extent of these mismatches, evaluate the need for matching algorithms to reconcile these differences. We also explore the feasibility of exploring the lead / lag relationship at more aggregated hierarchies (like aggregated at state / region | brand).

The final forecast specification will vary based on the insight derived through analysis of lead / lag relationships between the claim-dispense-data, sellout (867), and other data sources. However, based on current GTN forecast processes, we understand that there are several minimum requirement.

- Forecast Targets: Model will predict Medicaid utilization (units dispensed) per brand at national level. Will include every active Novartis Brand with a least 2 years of utilization history.
- Hierarchies: Model will predict total U.S. Medicaid units dispensed per brand, at minimum.
- Prediction Horizon: Model will predict utilization for ~4 months ahead of reported utilization based on Medicaid reporting. We will determine the optimal unit of prediction (e.g., week, month) during discovery and prototyping.[6]

We understand that the central finance team uses data aggregated by brand to national level and that will drive our explicit requirement. However, we will explore lead-lag relationships by state and more granular hierarchies, and see how this relationship varies based on different targets, like units dispensed, units billed, etc.,

Additional scope

Pending timing and success of 867 sellout data, we will also explore the relationship between dispense data and syndicated administration data and market-level events. (e.g., loss of exclusivity "LOE", formulary changes, etc.,)

In the first phase, we will primarily focus on predicting units (as of dispense or billed date) not price per unit. During the first phase of the engagement, we will understand the process for estimating unit rebate rate and following the completion of the unit forecasting model, we will work with Novartis to put this model into production in conjunction with current (or improved) process for unit-price estimation.

Scope 1A | Goals

1. **Provide a reusable and more accurate method for estimating Medicaid utilization during reporting gap**

Improve on existing estimates of Medicaid utilization (units dispensed) during reporting gap mapping relationships between claim-level-dispense data and relevant leading indicators (i.e., 867)

The MVP model deployed on Novartis infrastructure should be easily executable based on receipt of new data.

2. **Demonstrate that method can improve Medicaid gross-to-net forecast accuracy – i.e., full year sales estimate**

Produce analysis that demonstrates how model can be used during reporting cycles to improve central finance team total year sales forecast for Medicaid and improve their ability to hit their KPIs – 1% variance in out of -period adjustment.

Scope 2 | GTN Forecasting | Sequencing & Milestones

Project Stage	Development Milestones	Milestone Details	Target Date

[6] Otherwise referred to as the model "span".

Phase 1A	**[2.1]:** **Present Data & Modelling Insights**	Insights (PPT & Word) from analyzing relationship between claim-level-data, sellout data, and other data sources. Keystone to describe to Novartis how insights drive modelling approach.	Weeks 1 – 8
	[2.2]: **Executable MVP Forecasting Model**	**Medicaid utilization forecasting model, deployed in notebook on Novartis environment.** Model meets all "specifications" outlined above. <u>Execution:</u> Model code (training / inference) executable as a notebook on Novartis environment along with any data preparation code.[7] <u>Data Inputs / Outputs:</u> Data sources ingested into Keystone RAIN data platform and model outputs and artifacts are accessible via RAIN data platform.	Weeks 6 – 14
	[2.3] **Back Test MVP Forecasting Model**	Conduct an out-of-sample rolling backtest to quantify accuracy of utilization model across brands (MAPE/WAPE). Compare to previous estimates.	Weeks 10 – 14
	[2.4] **GTN Impact Analysis**	Produce analysis that demonstrates how model can be used during reporting cycles to improve central finance team total year sales forecast for Medicaid and improve their ability to hit their KPIs – 1% variance in out-of -period adjustment.	Weeks 14 – 16
Phase 1B	**Production Forecasting Model**	Exact scope TBD. Would include following HL items. 1. Add price estimation to model framework. 2. Analyze additional casual variables. 3. "Productionalize" model. Deploy code as dockized executable containers in model pipeline. Implement Model orchestration and CI/CD in Git repositories.	Weeks 16 – 28

[7] During this stage, data ingestion / cleansing code, may not be separate from the model code (training / inference code).

Scope 3: Rebate Invoice Automation

Overview

Same overview from scope detail 2 applies to this section.

Scope 3 | Rebate Invoice Automation: Phase 1A Solution Detail

Problem

The Medicaid rebate processing team (5 FTE) is currently overburdened reviewing (approving / triaging) low value Medicaid invoices, without the ability to triage / prioritize review of problematic and high-value claims vs. non-problematic / low value claims.

- Novartis receives ~5,000 rebate invoices from reporting state Medicaid entities that must be approved / disputed within 37 days.
- Invoices frequently have errors that overestimate the value of total rebates Novartis owes. Based on current processes, Novartis disputes ~$50M dollars' worth of rebates every year, approximately 5% of overall rebate dollars requested.
- The current process to review each claim is entirely manual and time consuming, taking on minimum ~45 minutes to review one invoice and potentially up to several days for others. Novartis analysts check for 4 general types of errors depending on the granularity of utilization data received with each claim. These checks are for – (1) units-of-measure errors, (2) 340B duplicates, (3) duplicates within the claim, and (4) duplicates across different claims.
- Rebate dollar requested by each invoice greatly vary and currently Novartis team has no way of quickly triaging smaller claims. Over last year, ~80% of claims received were less than 500K in request rebates but all of these sub-500K rebates only represented 11% of rebates paid. (vs. 90% for large rebate requests)

Solution

Keystone proposes a 14–16-week engagement to analyze historical Novartis historical claims, deploy a multi-stage AI model to probabilistically identify errors in each rebate invoice. This model would extract invoice terms from unstructured rebate invoices to produce a recommendation that (1) the invoice has no known errors or (2) flag invoice to human review due to error. The initial model would focus on sub-500k claims and likely start by automating checks that do not reference claim-level-data. We would pursue analysis of CLD in phase 1B.[8]

Data discovery & model approach: Keystone will review historical invoices received by Novartis in Model N, rationale / causes for disputes / approval, and understand the share of invoices disputed based on each dispute rationale.[9] This will be used to inform the model Build.

Data ingestion / LLM term extraction: Keystone will deploy LLM powered term extraction to extract key terms from unstructured invoices to be used in the claim triaging model. This will include terms like date, claimant / reporting entity, NDC, UOM, ID codes, units dispensed, provider ID, etc.,

Claim triaging model: Deploy a model to probabilistically predict inaccurate rebate requests, flagging claims above a probability threshold and approving claims below a reasonable threshold. Thresholds would be determined via out-of-sample back-testing to simulate historical performance.

---

[8] Exact scope of checks we build into AI model should be determined following review of data. Based on 11/19 discussion, Keystone understands that today, analysts do not use claim-level-data to review invoices below 500K. Our goals for the first phase 1A are to accelerate claim review based on current specifications, while 1B we'll expand on this to improve methods for processing claims based on CLD. (such as 340B dup checks)

[9] We understand that the history of these disputes are detailed in Model N.

- **Training data:** To train the model, we would use historically labeled rebate requests – both approved and rejected, along with their rationale – to build a reference dataset linking claim details to their corresponding dispute or approval outcomes.
- **Model outputs:** The trained model will output a recommendation on whether the invoice is problematic based on (1) UOM checks, (2) duplicates within the claim, and (3) duplicates across different claims. Additionally, it will also assess likelihood of dispute resolution by the reporting entity and the time to resolution. This would enable the model to flag problematic claims that are most likely to be successfully disputed.

Items out-of-scope for phase 1, to-be-deployed in phase 2

Claim-level-data verification: Because the solution will prioritize processing sub 500K claims, and today no claim-level-data verification is conducted on sub-500K claims, we will not incorporate those checks into the model yet. The first model will prioritize automatically conducting the checks currently done on sub-500K claims. In the second phase, we will focus on improving the checks conducted, like by conducting automatic claim-level 340B checks against Medicaid exclusion file, checking claim-level trends against broader commercial trends, and expanding automatic claim review to larger invoices (above 500k).

System integration: The MVP claim processing model will be deployed on Keystone's designated cloud environment on Novartis infrastructure but will not include automatic integrations to push / pull data from Model N or other systems. We assume that the model will be executable, and outputs are accessible.

Phase 1A | Goals

1. **Substantially reduce time-to-review Medicaid invoices below 500K by providing a reusable model for accurately predicting / identifying common submission errors**

Based on scoping discussions, we understand on average it takes ~60 minutes to review a sub-500K claim for 3 types of checks – (1) UOM, (20 Duplicates within Invoice, (3) Duplicates across invoice. We will deploy a model that reduces the time to manually review these invoices by assessing them programmatically with a high degree of accuracy.

2. **Demonstrate that method can be productized to provide additional time-savings, expanded to review additional error types, and used for above 500K claims**

Scope 3 | Rebate Invoice Automation | Sequencing & Milestones

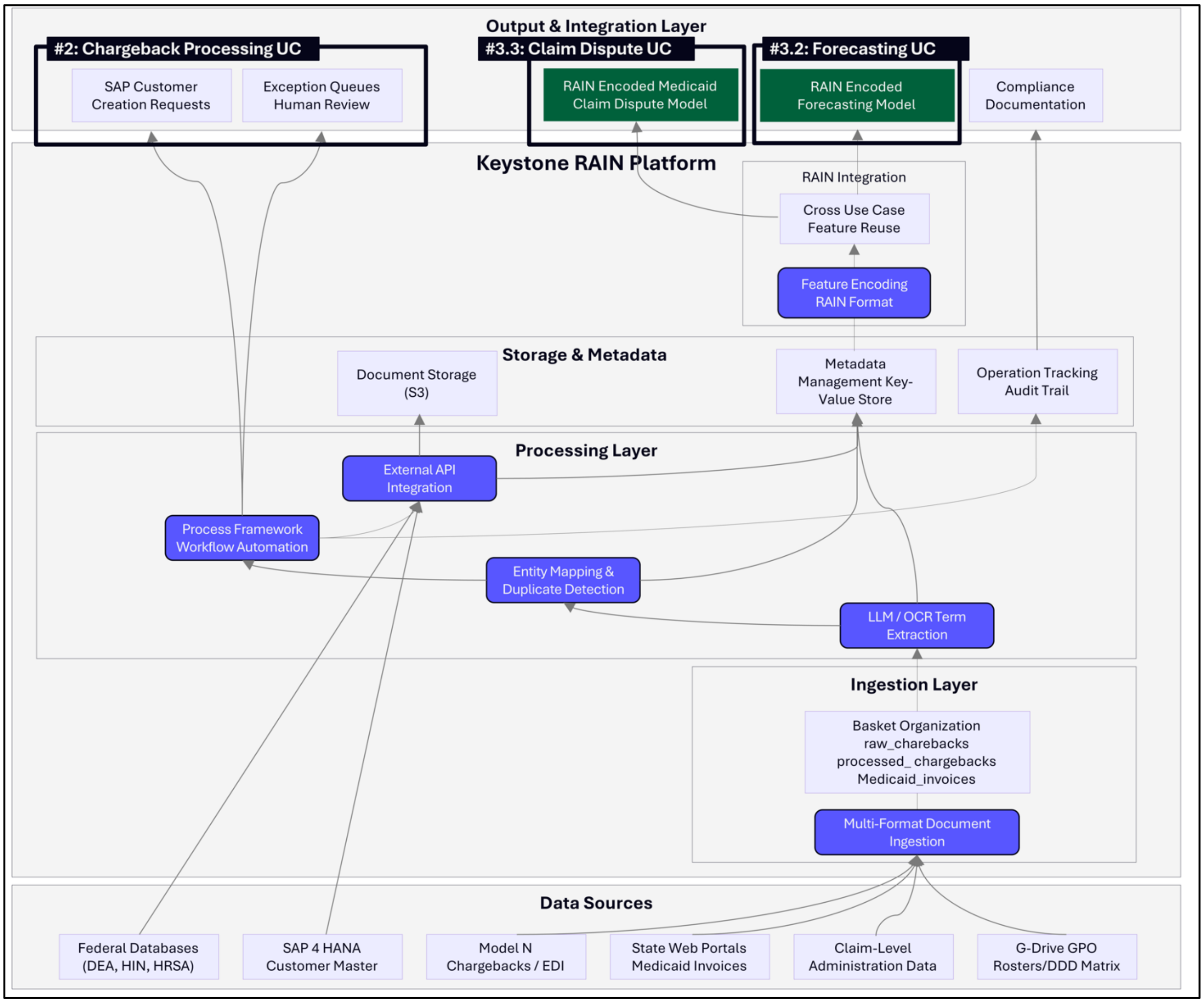
Project Stage	Development Milestones	Milestone Details	Target Date
Phase 1	[3.1] Claim Automation Data & Modelling Insight	Insights (PPT & Word) from analyzing historical rebate invoices (e.g., dispute volume / dollar-value broken down by type). Keystone to describe to Novartis how insights drive modelling approach.	Weeks 1 – 8
	[3.2] LLM Term Extraction Workflow	Functioning workflow (executable code) deployed on Novartis environment that extract key terms from unstructured rebate invoices and prepares them for use in the rebate invoicing model.	Week 10
	[3.3] MVP Rebate Invoice Triaging Model	MVP claim processing model, deployed and executable on Novartis environment, that take rebate invoice as input and outputs probabilistic recommendation to dispute or approve based on most-common dispute causes.	Week 14

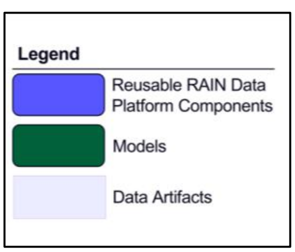
	[3.4] Model Back Test and Summary of Model Performance	Conduct an out-of-sample backtest to quantify accuracy of claim processing model (% error) using historical sub-500K claims not used for model training. (i.e., out-of-sample)	Week 16
Phase 1B	Production Rebate Invoice Triaging Model	Exact scope TBD. Would include following HL items. 1. Productuctize existing model in reusable workflow for Novartis. Source code deployed as dockized containers on Novartis infrastructure. 2. Deploy automatic integrations to Model N. 3. Expand model capabilities to do claim-level-data verification, such as 340B Duplicate Check, channel trend identification. 4. Test and expand model coverage to above 500K invoices.	Weeks 16 – 28

Exhibit: Solution Architecture

RAIN Data Platform Overview

Figure 1: Keystone RAIN Platform Key Components & Relationships





This high-level diagram shows the solution components, how they connect, and how they work together across use cases. The following sections describe how each “Core Solution Component” (outlined in purple and green) is specifically used for each use case.

Data / System Dependencies

Additional Process Detail

Chargeback Processing: Process Context & Relevant Details from 11/19 Scoping Call

1. **Current customer creation process is “reactive” not “proactive”.** Novartis has up-to-date details on each customer contract in the G-Drive document and *could* create new customer IDs in SAP 4 HANA preemptively but chooses to wait until that customer has been submitted in a chargeback EDI because this is less time intensive than creating new customers in SAP based on every contract in the roster. (i.e., there are 1000s of customers on the GPO Roster that will

never purchase a Novartis product) Key Deployment Implication: Right now, our intent is to automate a reactive workflow, in the long-term eliminating this workflow entirely would be possible through a proactive process to create customers intelligently.

2. **99.9% of the time, a customer is created based on this error.** This is because this error is expected. Novartis doesn't create new customers until the chargeback EDI is submitted for that customer (i.e., reactive process). In the other 0.1% of cases, the customer is not eligible on the submitted contract (i.e., steps 4&5) then the chargeback would be rejected / unresolvable. Key Deployment Implication: A perfect system would create customer requests for 99.9% of chargebacks with this error received but also account for the 1% of cases where the customer doesn't exist.
3. **"Customer not found" kickouts represent ~15% of kickout types**, but they are the most manually time intensive to resolve. Other "kickout" types include (1) product is not found on EDI, (2) wholesaler submitted the incorrect contract, (3) wholesaler submitted incorrect price on EDI. Key Deployment Implication: Long-term goal is to eliminate manual triaging for those kickout as well.
4. **"Customer not found" kickouts explicitly mean** that all "ship to" identifiers submitted by the wholesaler are not found in the system. A customer can submit up to 3 identifiers per each chargeback line and Model N will validate the chargeback line of each identifier against the system – SAP 4 HANA.
5. **G-Drive "Knowledge Base" references** including: (1) Membership Eligibility Guide, (2) GPO Roster, (3) DDD Matrix. These are updated every time a new contract / contract amendments are received and serves as a user guide across each step for membership analysts. These references are also used to calculate rebate payments. Key Deployment Implication: While not a goal to start, automating how this is updated represents long-term goal.
6. **"Class of Trade" selection requires discretion.** COT can vary based on lower-level customer distinctions like the type of hospital or type of clinic. Fed website information is used in conjunction with the DDD matrix to make this selection.
7. **Most "customer not found" kickouts are from commercial customers not Medicaid or 340B providers.** Those customers would be pre-approved. Solution Implication: DEA and HIN are critical Fed websites.
8. **Novartis experience with API calls to DEA and HIN has faced issues with data lags.**
9. **Several screenshots are taken for each customer** created for tracking & compliance in a Word doc attached in SAP for the customer. These are (1) relevant Fed website, (2) class-of-trade (presumably DDD matrix), (3) collaboration ticket in Model N.
10. **Priorities from Novartis are API calls to Fed websites then GPO Roster pulls.**