

Workflow Documentation

Phone & Email Enrichment and Deduplication Pipelines

Internal Operations — Automation Team

1. Purpose & Scope

This document provides a technical overview of two automated workflows developed to enrich and validate contact data — specifically phone numbers and email addresses — and to maintain data integrity across overlapping contact sheets. These workflows form a critical part of the outreach data pipeline, ensuring that contact records are complete, verified, and free of duplication before any downstream processing.

The two workflows covered in this document are:

- Workflow 1 — Multi-API Contact Enrichment & Verification Pipeline
- Workflow 2 — CEO / CEOI Sheet Deduplication Workflow

2. Workflow 1: Multi-API Contact Enrichment & Verification

2.1 Overview

This workflow is designed to enrich contact records with missing phone numbers and email addresses, and to verify the accuracy of those already present. It operates on a waterfall (sequential fallback) model across multiple third-party enrichment APIs, ensuring maximum data coverage even when individual API calls fail or return no results.

2.2 How the Waterfall Model Works

The waterfall approach means that each enrichment step is attempted in order. If the first API successfully returns a valid result, the workflow proceeds without calling the next API. If it fails or returns no data, the workflow automatically falls through to the next API in the sequence. This chain continues until either a result is obtained or all available APIs have been exhausted.

This design provides three key advantages:

- Redundancy — No single point of failure can leave a record unenriched.
- Cost efficiency — More expensive or rate-limited APIs are only called when cheaper ones fail.

- Coverage — Using three separate enrichment APIs significantly increases the probability of retrieving valid contact data.

2.3 Enrichment APIs

A total of three enrichment APIs are integrated into this pipeline. Their exact identities may vary by deployment, but their roles in the waterfall are fixed:

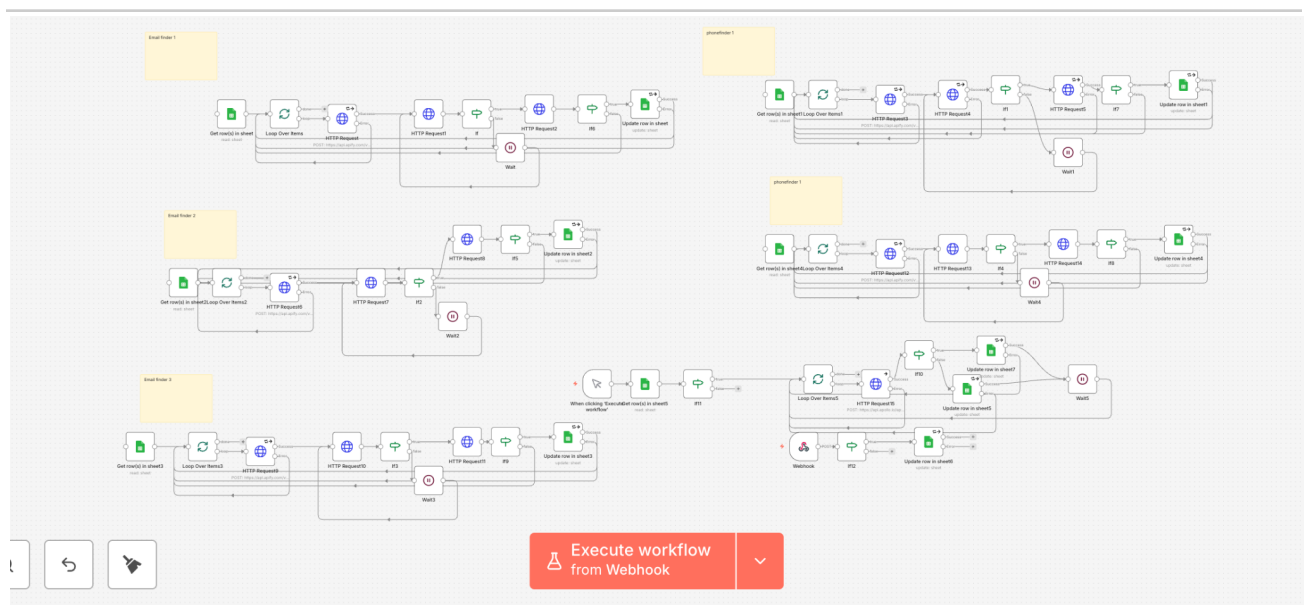
- API 1 (Primary) — First attempt for both phone and email enrichment.
- API 2 (Secondary) — Called only if API 1 returns no result or an error.
- API 3 (Tertiary) — Final fallback, used when both prior APIs have failed.

2.4 Verification Step

In addition to enrichment, the workflow includes a verification step that validates the enriched (or pre-existing) contact data. This ensures that phone numbers are dialable and correctly formatted, and that email addresses are syntactically valid and not associated with known bounce domains. Verification runs after enrichment and flags or removes records that do not pass quality checks.

2.5 Workflow Diagram

The diagram below illustrates the full enrichment and verification pipeline, including the fallback sequence across all three APIs:



3. Workflow 2: CEO / CEOI Sheet Deduplication

3.1 Overview

This workflow addresses the problem of duplicate contact records appearing across two separate sheets: the CEO Sheet and the CEOI Sheet. Before the enrichment pipeline (Workflow 1) is executed, this deduplication step ensures that any contact already present in the CEOI Sheet is removed from the CEO Sheet. This prevents the same individual from being processed and enriched twice, saving API credits and keeping the data clean.

3.2 Deduplication Logic

The workflow performs a cross-sheet comparison using a defined set of matching criteria (such as name, company, or a unique identifier). The logic proceeds as follows:

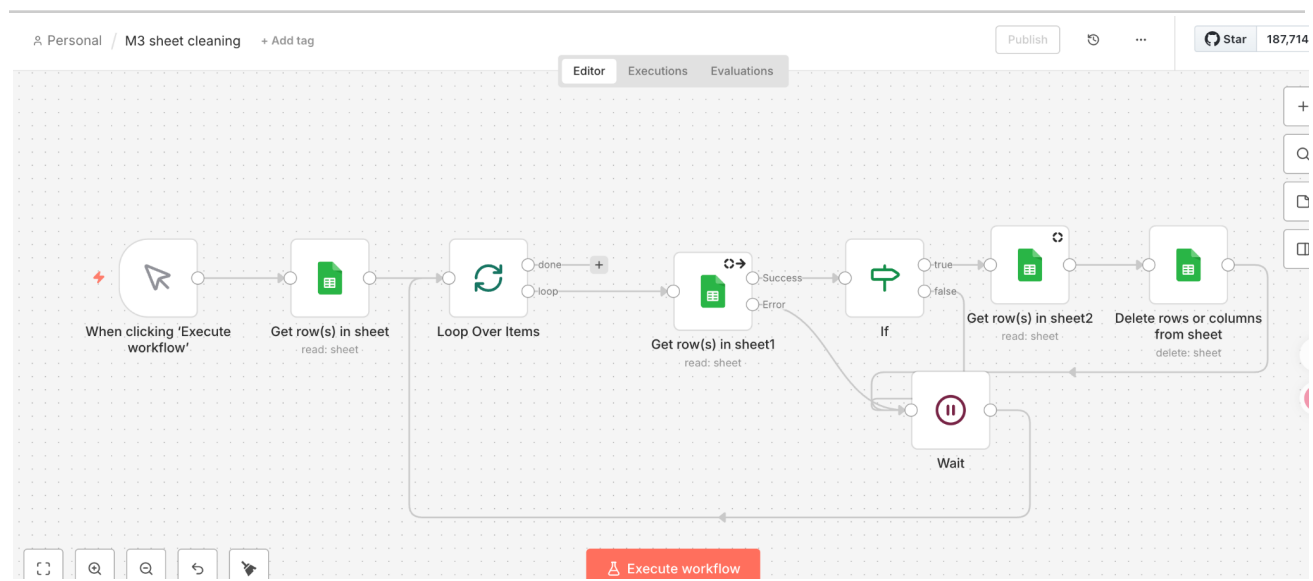
- The CEOI Sheet is loaded as the reference dataset.
- Each record in the CEO Sheet is checked against the CEOI Sheet.
- Any CEO Sheet record that matches an existing CEOI Sheet entry is flagged as a duplicate.
- Flagged duplicates are removed from the CEO Sheet before enrichment begins.

3.3 Why This Step Matters

Running enrichment on duplicate records wastes API calls and can introduce conflicting or inconsistent data into the final output. By performing deduplication upfront, this workflow guarantees that the enrichment pipeline only processes unique, non-overlapping records — improving both data quality and operational efficiency.

3.4 Workflow Diagram

The diagram below illustrates the cross-sheet comparison and deduplication process:



4. End-to-End Pipeline Summary

Together, these two workflows form a complete contact data preparation pipeline. The recommended execution order is as follows:

Step 1	Run the Deduplication Workflow (Workflow 2) to clean the CEO Sheet by removing any records already present in the CEOI Sheet.
Step 2	Run the Multi-API Enrichment & Verification Workflow (Workflow 1) on the cleaned CEO Sheet to populate missing phone and email data, and verify existing contact information.

Executing these in order ensures maximum efficiency: enrichment is only performed on records that are both unique and have been confirmed as not already captured in the CEOI dataset.

5. Key Design Principles

- Fault tolerance — Waterfall API chaining prevents data gaps caused by any single provider outage.
- Data integrity — Deduplication upstream ensures clean inputs to all enrichment steps.
- Verification-first mindset — All enriched data is validated before being written to the final output.
- Scalability — The modular structure allows new enrichment APIs to be added to the waterfall chain with minimal rework.

6. Notes & Maintenance

The following operational notes apply to both workflows:

- API credentials must be kept up to date. If any enrichment API key expires, the corresponding step will silently fail and the fallback will be triggered.
- The matching criteria used in the deduplication step should be reviewed periodically to ensure they remain accurate as data schemas evolve.
- Verification logic should be updated if the acceptable email domain or phone format rules change.

For questions, modifications, or onboarding new team members, refer to the internal automation runbook or contact the responsible workflow owner.