Design Note: Credit Card Data Harvester

The Credit Card Data Harvester is a structured n8n-based workflow designed to extract and consolidate credit card product data from two sources: official MITC/KFS PDFs and public product webpages. Upon form submission, the system fetches HTML content from the URL and extracts text from the uploaded PDF.

Two Gemini-powered AI agents extract 29 standardized fields from the PDF and HTML respectively. Citations are included with each extracted value (page numbers or CSS selectors). A Code node then merges these outputs using rule-based logic: PDF values are prioritized; if missing or invalid, HTML values are used; otherwise, it defaults to 'Not mentioned'. The node also performs normalization (e.g., combining tag arrays).

The final merged output includes all 29 fields, ready for conversion or email delivery. This AI-assisted approach balances flexibility and traceability through structured outputs and embedded citations.

Proposed Improvement: With more time, a semantic validation and reconciliation layer can enhance reliability. This includes entity recognition, terminology standardization (e.g., 'Annual Fee' vs. 'Renewal Fee'), and unit normalization. Semantic comparisons would flag inconsistencies between sources using embeddings and natural language understanding. Confidence scores can be introduced per field to guide human review for ambiguous or conflicting data. A web-based interface would enable manual correction, approval, and export.

Such enhancements would elevate the workflow from rule-based merging to intelligent data synthesis—making it more adaptive, reliable, and production-ready.