

AI Attribute Intelligence

Automating Detection, Extraction, and Standardization at Scale



Melissa Hollis

Director - Data Analytics



Zach Dygert

Sr Data Scientist



Justin Claycomb

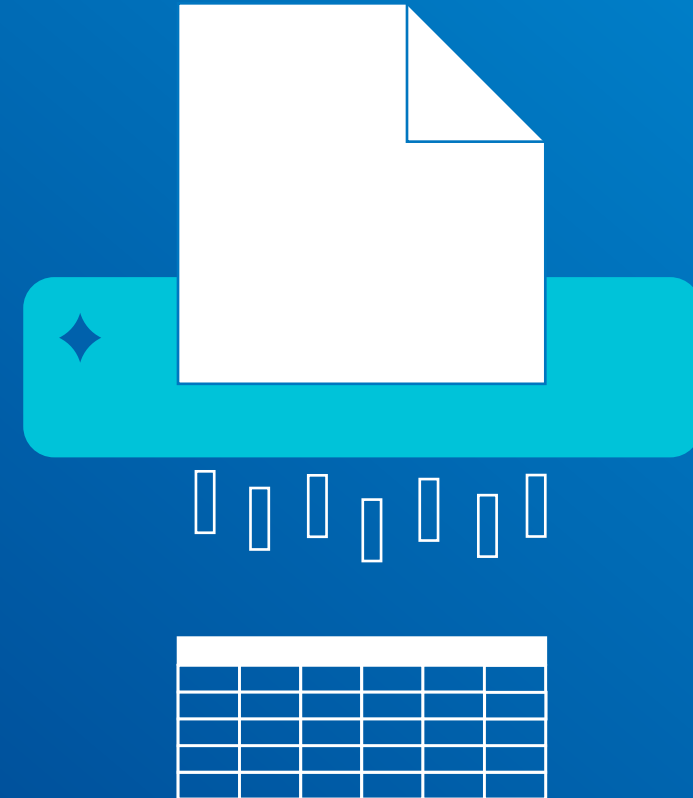
Asst. Dir - Data Science

Agenda

Setting the stage

Solution walkthrough

Value of scaling



Setting the stage

When Humans Are the Interface

Food Delivery App

Every option is constrained:

- You can only choose what's on the menu
- You can only enter values in the format the system expects
- If something doesn't exist, you simply can't submit it

The system does the thinking up front.

Download -> Print -> Hand write -> Fax

A human on the receiving end must:

- Read it
- Interpret what you meant
- Match it to the menu
- Correct inconsistencies
- Re-enter it into their system

Most Enterprise Data Arrives Like a Fax

Most data entering enterprise systems looks a lot more like the fax than Food Delivery Apps

Not because people are careless - but because:

- Inputs come from many external parties
- Requirements vary by context
- Standards aren't enforced at the point of creation

So instead of clean, structured, system-ready inputs, organizations receive:

- Spreadsheets
- Forms
- PDFs
- Emails
- Free-text messages
- Files that are *close* to what's needed, but not quite

And just like the faxed order, a human has to interpret it before a system can use it.

Opportunity for Augmentation

This creates a pattern most organizations recognize:

- Data arrives in inconsistent formats
- Field names vary
- Values don't line up
- Context lives outside the file

So teams step in to:

- Read
- Translate
- Reformat
- Map
- Re-enter

That work is:

- Manual
- Time-consuming
- Repetitive
- Error-prone

And critically - **it does not scale**

Opportunity for Augmentation

When humans are responsible for interpretation

- Onboarding new data sources takes longer than it should
- Downstream systems wait on upstream cleanup
- Changes require rework instead of reuse
- Knowledge lives in people, not in processes

The result isn't just inefficiency - it is **delay**

Delay in

- Activating services
- Delivering value
- Gaining insight
- Building trust in the data

From the outside, it looks like a processing delay.
Inside the organization, it's really an interpretation delay

This Is Not an Industry Problem

Any organization that:

- Accepts data from outside its walls
- Works across departments or partners
- Relies on attributes, fields, or metadata for automation

Faces the same challenge.

The more data you ingest, the more interpretation you inherit.

**Why are humans still acting as translators
between messy inputs and rigid systems?**

Solution

Juggling the Chaos

There is diversity across dimensions

It is not just an unstructured data problem

File types

Structure

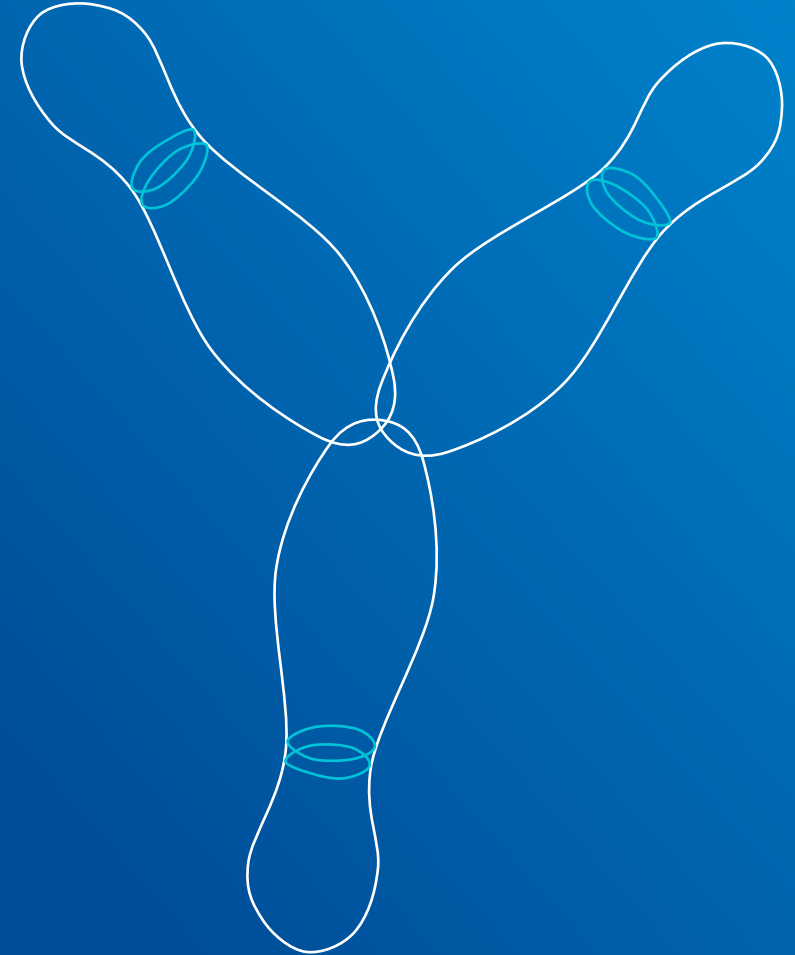
Formats

Field names

Overloaded values

Context

... Data itself



Intake Flexibility

File type guides

Various modules depending on type

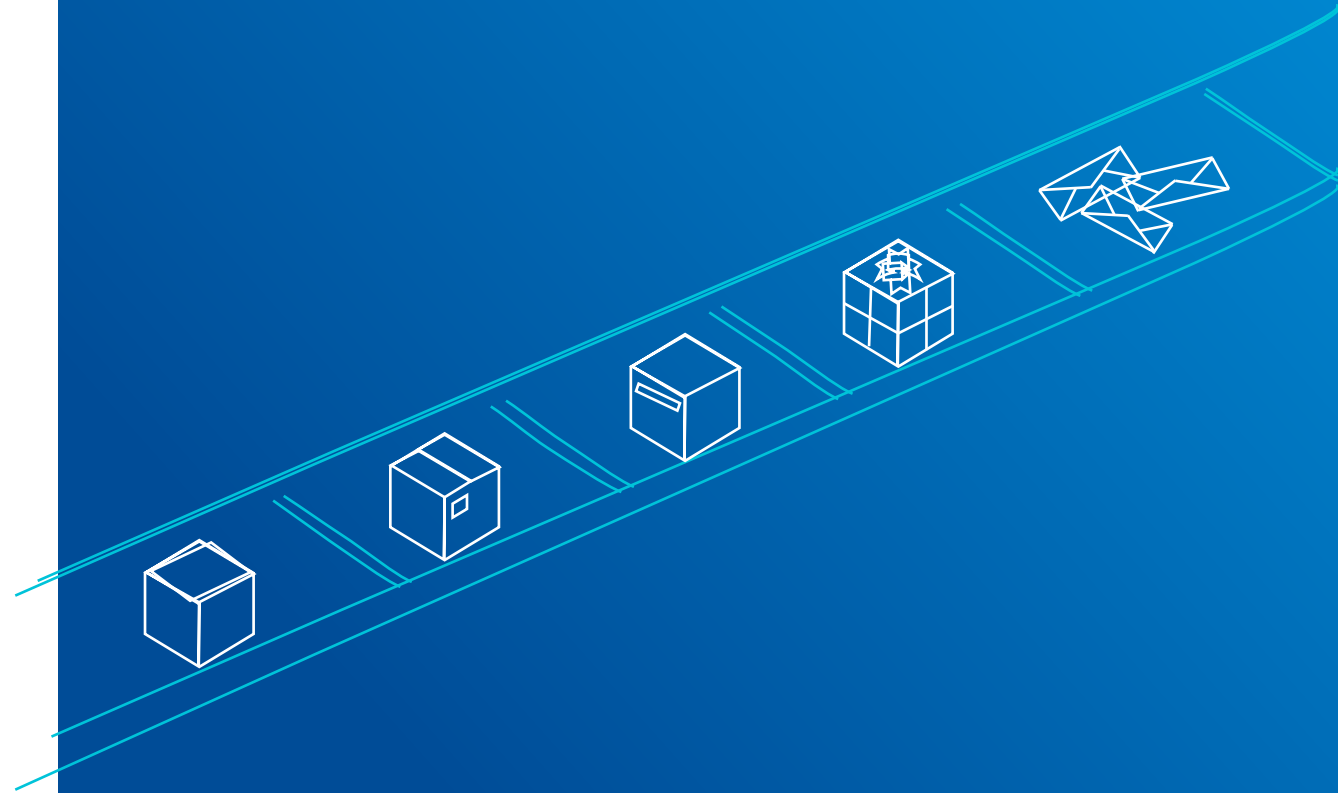
`check_file_type()`

`read_file()`

`structure_cleanup()`

...

`extraction()`



Tiered Engine

Increases efficiency and reduces risk

Simple to Sophisticated Stack

Document and Text Operations

Basic Natural Language Processing

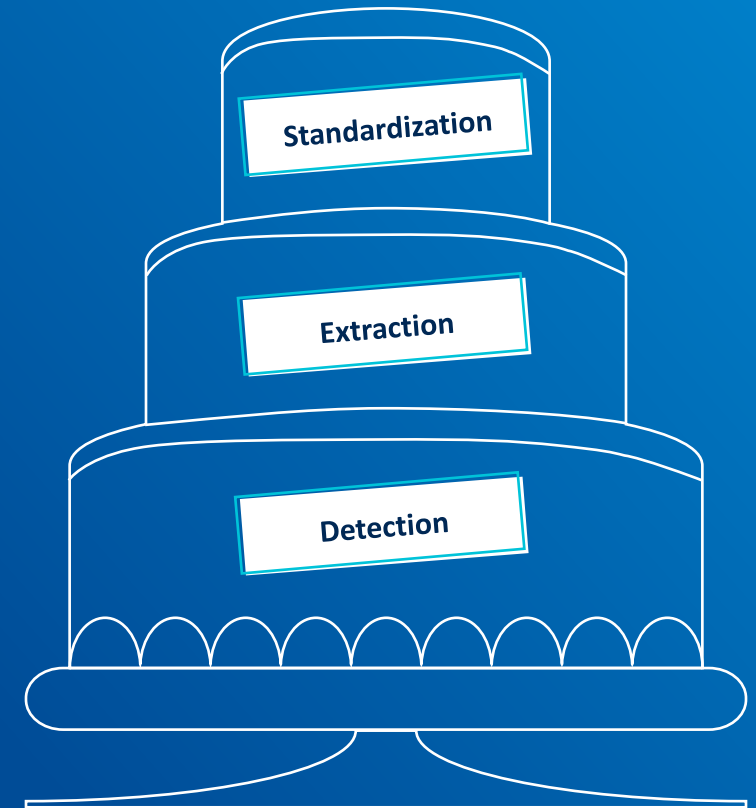
Machine Learning

Named Entity Recognition

Large Language Models/ RAG

Calculations

Data Validation



Double Click

Dictionary Intelligence

Domain/ business knowledge guides

Curate synonyms and feedback loops

Trackable/ auditable

Higher precision

Semantic ML

Learns from dictionaries

Allows for variation and contextual queues

Embeddings + voting

Probabilistic control

Confined scope

LLMs/ RAG

Parsing/ section isolation/ chunking strategy

Model/ parameter selection

Prompt engineering/ instructions

Data validation steps

Important Considerations

Human Expertise

Human in the loop

Validation and quality checks

Defined compliance and risk boundaries

Data Privacy

Data minimization and purpose limitation

Redaction or filtering of sensitive data

Retention and access controls

Consent and appropriate use boundaries

Safety Controls

Risk assessments

Validation checkpoints

Alignment to enterprise governance standards

Scaling

This is a pattern, not a one-off

From Use case to Capability

- Moving from individual use cases to a shared attribute intelligence platform
- Standardizing a common workflow
- Reducing one-off build patterns through reusable component
- Shifting team effort away from mechanics and toward data quality, confidence, and reuse

Make Adding New Attributes Fast

- We add new attributes through configuration, not custom builds.
- Every attribute follows the same clear definition pattern.
- Rules first, AI only when it's needed.
- The same engine works for structured and free-text data.

Speed without Sacrificing Control

- New Data Sources follow a repeatable onboarding path
- Existing pipelines reused, differences handled through configuration
- Minimal new engineering for most expansions
- Built-in validation and confidence testing with outputs
- Enables faster adoption while maintaining trust and consistency

Questions



Melissa Hollis
Director - Data Analytics



Zach Dygert
Sr Data Scientist



Justin Claycomb
Asst. Dir - Data Science