CHAPTER 3: CASTING THE NETS | TRY THIS NOW

Know Your Data

Find the baggage in your data before you automate it.

Two quick ones from Chapter 3. Data shows up with baggage. These help you see it before you automate anything.

TRY 1 Who decided what your tool can see?

Look at a database or tool you use regularly.

Database or tool

What fields or categories does it force you to use?

What can't you track because there is no field for it?

Who decided those categories were important, and what might they use them for that is different from what you need?

TRY 2 Sort your recent data

Grab data you collected recently: a report, imagery, social media posts, anything. Sort each item. Note anything that doesn't fit neatly in one box.

Item	Structured	Semi-structured	Unstructured	Doesn't fit cleanly? Why?

Mark the column with an X. Structured = rows and columns with fixed rules. Semi-structured = tags and metadata, like JSON or API feeds. Unstructured = posts, imagery, audio, PDFs.

What surprised you?

BONUS Provenance spot check (Table 3.1)

Pick one random data point from your workflow and see if you can answer all three.

Provenance question	My answer for this data point	Can I answer it?
Origin: where did this come from? (URL, timestamp, collection method)		
Transformation: what happened to it? (cleaning, renaming, AI preprocessing)		
Custody: who handled it? (tools, versions, models, human checkpoints)		

If you can't fill in that last table, you don't have provenance, and you can't lean on analysis built from that data.