30th Annual NSPAA Technical Assistance Workshop | June 2026

LLM-Assisted Data Mapping for Sponsored Programs

A Human-in-the-Loop Data Cleanup Workflow

Mana Rassaeikashuk, MSc, MBA | Doctoral Student | Lead Presenter

Dr. Esther Olaleye, PhD, MPH | Senior Associate Vice President for Research and Innovation | Professor of Pharmacology and Epidemiology | Co-presenter

TSU Texas Southern University Division of Research & Innovation | Texas Southern University Joan M. Lafleur College of Pharmacy & Health Sciences | CBMHR Center for Biomedical & Minority Health Research Community Engagement Core

PROUD TO BE A Carnegie R2 and Community Engagement University

TSU Texas Southern University

NEAT

No-code Entry Alignment Toolkit

TSU NEAT is a Cleanup Assistant for Sponsored Programs Data; It Suggests, Explains, and the Human Approves.

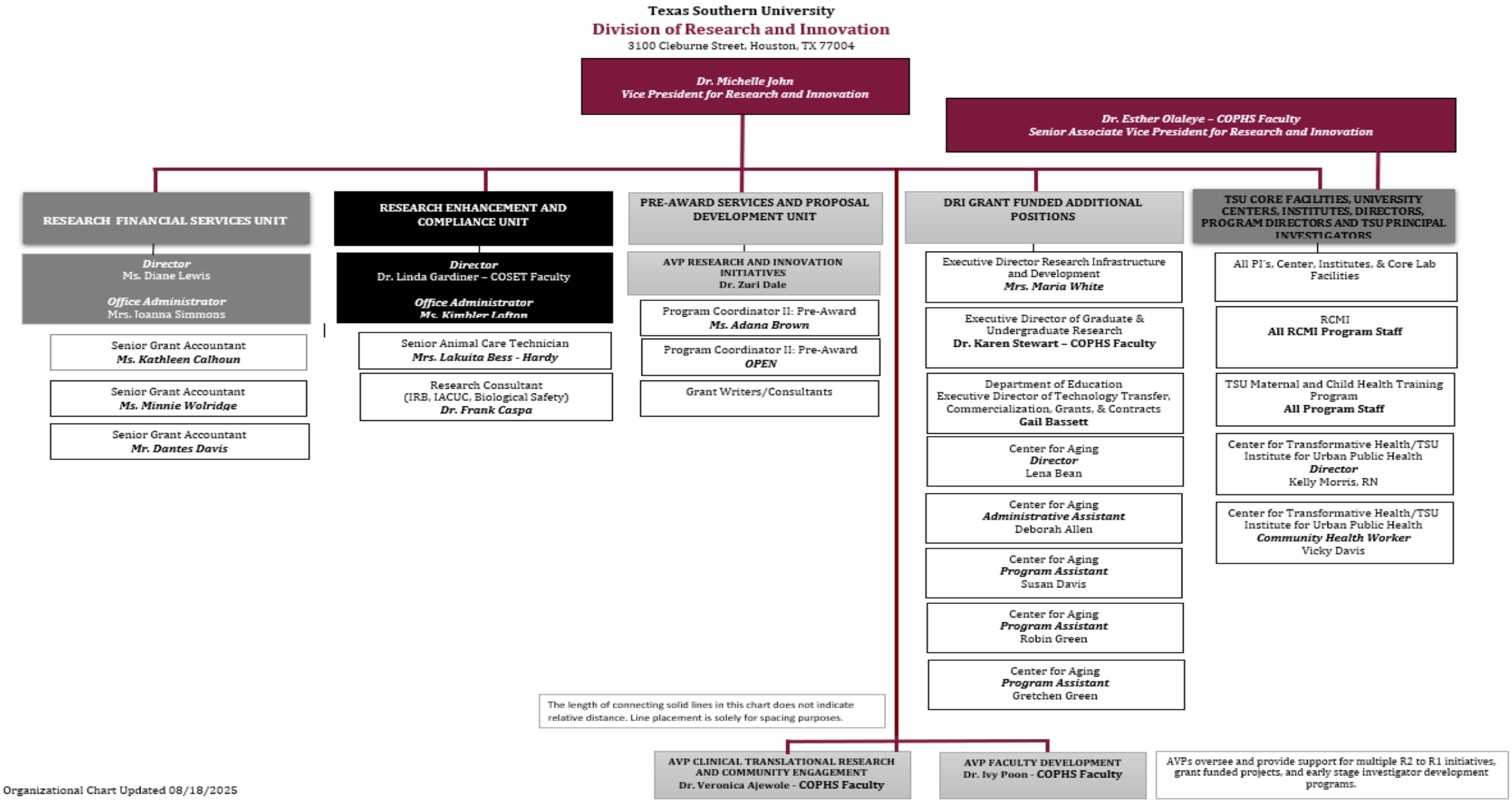
LEARNING OBJECTIVES

1. Apply a step-by-step, human-in-the-loop workflow to standardize one sponsored program's free-text field using a controlled vocabulary and model assistance.

2. Implement three governance safeguards (data minimization, de-identification, and reviewer audit logs) to reduce risk and support compliance.

3. Create a pilot plan with success metrics (accuracy, reviewer time, consistency) and a reusable mapping dictionary.

Division of Research & Innovation — Organizational Structure



With acknowledgment to Dr. Michelle John and Dr. Zuri Dale (Texas Southern University).

Centers & Institutes — Overview

Texas Southern University houses 16 research centers and institutes spanning health sciences, transportation, computing, law, and community development.

Biomolecular Research and Advanced Computing Centre
Computational Sciences

Center for Justice Research
Justice & Policy

Center for Biomedical and Health Research
Biomedical / RCMI

High Performance Computing Center
Computing

Center for Transformative Health
Public Health

Institute of Urban Public Health and Housing
Public Health

Gulf Coast Consortium Center for Comprehensive PK/PD and Formulation
Cancer Research

Center of Excellence for Housing and Community Development Policy Research
Housing

Earl Carl Institute
Legal & Social Policy

Innovative Transportation Research Institute
Transportation

Gulf Coast Collaborative Center for Maternal Health Research
Maternal Health

Advance Health Program
Public Health

Center of Aging and Intergenerational Resources
Aging & Community

Center for Transportation Training and Research
Transportation

Forensic Sciences Learning Lab
Forensic Science

Institute of Drug Discovery & Development
Drug Discovery

With acknowledgment to Dr. Michelle John and Dr. Zuri Dale (Texas Southern University).

Carnegie Community Engagement Classification

The gold standard for community engagement in U.S. higher education.

CARNEGIE COMMUNITY ENGAGEMENT CLASSIFICATION

4,000+	< 400	Top 10%
U.S. Degree-Granting Institutions	Hold This Classification	of All U.S. Colleges

Fewer than 1 in 10 U.S. institutions hold this designation

WHAT CARNEGIE EVALUATES

- Reciprocal Partnerships
- Community-Engaged Research
- Institutional Leadership
- Sustained Assessment

WHY IT MATTERS FOR TSU

Funders
Competitive advantage for community-engaged research grants

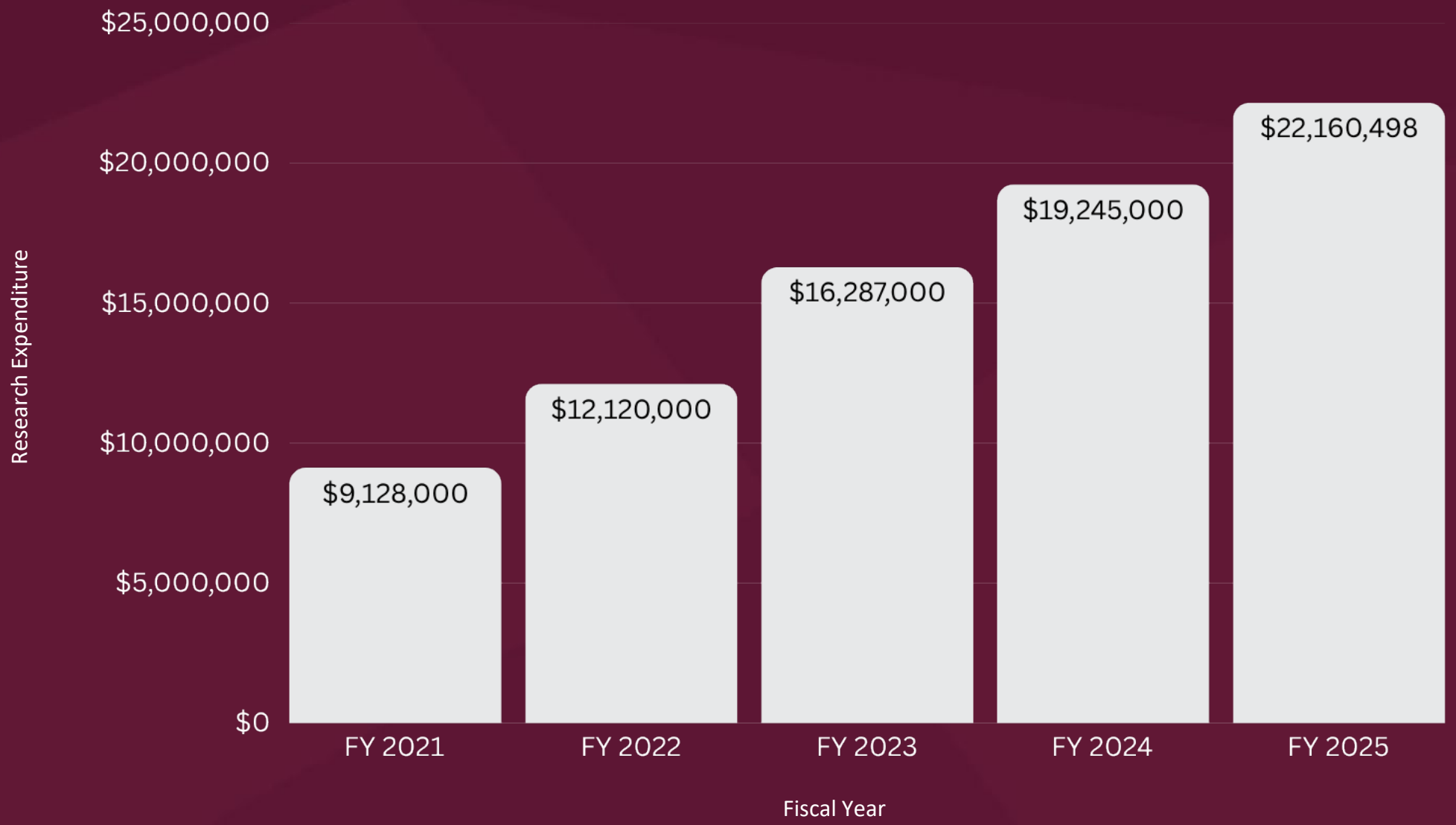
Students
Proof of real-world, mission-aligned education

Partners
TSU is a trusted, accountable reciprocal partner

Peers
Rare dual distinction: R2 Research + Carnegie combined

With acknowledgment to Dr. Michelle John and Dr. Zuri Dale (Texas Southern University).

Research Expenditures by Fiscal Year



With acknowledgment to Dr. Michelle John and Dr. Zuri Dale (Texas Southern University).

THE “5-MINUTE ROW” IS REALLY A CAPACITY PROBLEM

“69% of respondents reported they lacked the right staffing resources.”

Cayuse, 2022 Benchmark Report, cited in Cayuse, 2025a

“RAs have to clean up the information and make it more digestible, which takes even more time.

Cayuse. (2025b). *7 challenges with homegrown research administration solutions.* Cayuse.

Minute 1
Review inconsistent entry

Minutes 2–3
Identify the correct label

Minutes 4–5
Re-label and document correction

When senior research administrators stop losing hours normalizing “NIH” versus “National Institutes of Health,” those hours can be redirected toward portfolio analysis, faculty support, and grant development.

Institutions cannot build toward R1-level research capacity when high-skill staff are bogged down doing free-text spreadsheet cleanup.

THE HIDDEN CAPACITY DRAIN

Messy free-text fields quietly consume staff time.

What staff see

- Sponsor names entered 10 different ways
- Categories mixed in free text
- Every report cycle: cleanup again

Result: time spent normalizing instead of supporting strategy.

What leadership feels

- Slower reporting
- Inconsistent counts
- Harder audit response
- Fewer hours for high-value work

Example: sponsor name cleanup

- NIH
- Natl Inst of Health
- N.I.H. / NCI
- National Institute of Health

→ **Federal Sponsor**

Standard category

THINK OF NEAT AS A MAILROOM SORTING SYSTEM.

Messy input

"NIH"
"N.I.H."
"Natl Inst of Health"
"Smith Family Trust"
"Project X"

Federal | **State/Local** | **Foundation** | **Industry** | **Internal**

The assistant

Acts like a highly constrained intern: it suggests a bin and explains why.

The manager

A human reviewer makes the final call, logs the decision, and becomes the official system of record.

The asset

Every verified decision goes into a mapping dictionary, so the institution does not start from zero next time.

NEAT is not "IT software." It is a governed administrative workflow that helps research administration team sort and document decisions.

DATA GOVERNANCE

Clear boundaries make the method usable in real offices.

Data Hygiene

- ✓ De-identified or synthetic demo data
- ✓ Minimum necessary fields
- ✓ No award numbers, no PI names

Hallucination Control

- ✓ Controlled vocabulary only
- ✓ Structured output
- ✓ “Other” + ow confidence when ambiguous

Accountability

- ✓ Human verification is documented
- ✓ Review log optional but recommended
- ✓ Dictionary entries are verified

WHAT NEAT IS (AND IS NOT)

NEAT is Decision support — not automation

NEAT is…

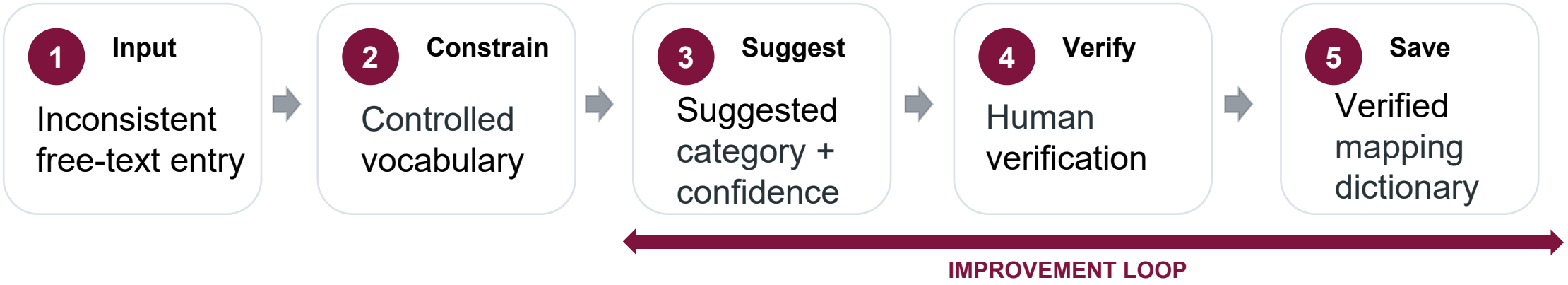
- ✓ A no-code workflow
- ✓ A controlled vocabulary (the safety rail)
- ✓ Model suggestions with confidence + rationale
- ✓ Human approval as the final decision
- ✓ A mapping dictionary that improves over time

NEAT is not…

- X A black-box “auto-cleaner”
- X A replacement for staff judgment
- X A system that invents categories
- X A tool that uses sensitive identifiers or live institutional data in demos
- X A one-off prompt with no governance

The model suggests. The human decides. The dictionary remembers.

THE GOVERNED WORKFLOW



What the institution keeps

A verified mapping dictionary + review notes (institutional memory).

Pattern	Category	Verified by / role	Date
NIH / NCI variants	Federal	Research admin reviewer	—
Gates Foundation variants	Foundation	Research admin reviewer	—
Pfizer / J&J	Industry	Research admin reviewer	—

AI drafts; humans decide; the institution keeps the asset.

LIVE DEMO

Excel Workbook (Toolkit)

Demo_Data | Controlled_Vocab | Prompt_Template | Mapping_Dictionary | Review_Log

Row	Sponsor Name Free Text	Suggestion	Confidence	Human
001	NIH	Federal	High	✓
016	Bill & Melinda Gates Fdn	Foundation	High	✓
020	Pfizer Inc.	Industry	High	✓
029	Community Partners Coalition	Other	Low	—

Constrained Prompt

ALLOWED CATEGORIES
[Federal, State/Local,
Foundation, Industry,
Internal, Other]

OUTPUT
row_id, category,
confidence, rationale

RULE
If unclear → Other + Low confidence

When the model says 'Other' or 'Unclear' with low confidence, that is a good result.

It means the workflow is not pretending to know what it does not know.

A PRACTICAL PILOT PATH

Start with one field. Measure value before expanding.

Recommended first pilot field: Sponsor Type

Weeks 1–2
Scope + hygiene

- ✓ Choose one field
- ✓ Confirm categories

Week 3
Constrained suggestions

- ✓ Run mapping
- ✓ Capture confidence + rationale

Week 4
Human review

- ✓ Correct edge cases
- ✓ Log final decisions

Weeks 5–6
Dictionary + rerun

- ✓ Save patterns
- ✓ Re-run and measure

Success metrics

Auto-accept Rate

- ✓ % accepted as-is
- ✓ **Target: >80%**

Speed / Throughput

- ✓ Reviewer time saved vs. manual cleanup
- ✓ **Target: major time savings**

Error Rate On High

- ✓ Spot-check “High” suggestions
- ✓ **Target: <1%**

Confidence Calibration

- ✓ Low-confidence rows should capture the hard cases
- ✓ **Goal: focus human time on the hardest 20%**

TWO HIGH-IMPACT USE CASES

Pre-award / Compliance

- ✓ Input: Sponsor name + description
- ✓ Output: HERD-aligned sponsor categories
- ✓ Value: cleaner reporting and less audit rework

Research Analytics

- ✓ Input: Project title and abstract
- ✓ Output: standardized research domains (Health, STEM education, Social Sciences)
- ✓ Value: portfolio analysis aligned to institutional priorities

The workflow stays the same.
That is why this matters beyond one spreadsheet.
It is a repeatable approach for organizing messy information while preserving human judgment.

COMPOUNDING TIME: CAPACITY FOR THE R2 → R1 TRANSITION

Micro Gain

A small gain at the row level can become a larger gain across many records.

Macro Shift

Across thousands of rows, that becomes dozens or hundreds of reclaimed staff hours.

Strategic Pivot

Those hours can be redirected toward advanced portfolio analysis, targeted faculty support, and grant development.

Institutional Momentum

Capacity you need to move from baseline compliance toward higher research activity.

Institutions cannot build R1-level research capacity when highly skilled staff are repeatedly pulled into spreadsheet normalization.

WHAT ATTENDEES CAN TAKE HOME

NEAT
No-code Entry Alignment Toolkit

You do not need a massive IT budget or a new enterprise system to pilot responsibly.

Workbook

Demo data
Controlled vocabulary
Mapping dictionary

Prompt Pack

Constrained prompts
Structured output template
Confidence + rationale

Governance Rules

Safe-use boundaries
Data minimization
Human review and final decision

Pilot Checklist

6-week pilot path
Success metrics
Documentation

NEAT is a method you can run now with familiar tools.

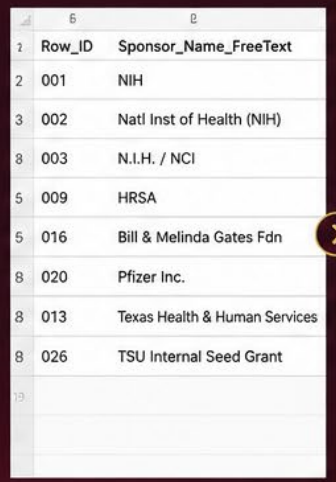
NEAT

No-code Entry Alignment Toolkit

AI suggests.
The reviewer decides.
The reviewed result becomes the system of record.

A HUMAN-IN-THE-LOOP WORKFLOW FOR DATA NORMALIZATION

1 Messy Input (Excel)



Row_ID	Sponsor_Name_FreeText
001	NIH
002	Natl Inst of Health (NIH)
003	N.I.H. / NCI
009	HRSA
016	Bill & Melinda Gates Fdn
020	Pfizer Inc.
013	Texas Health & Human Services
026	TSU Internal Seed Grant

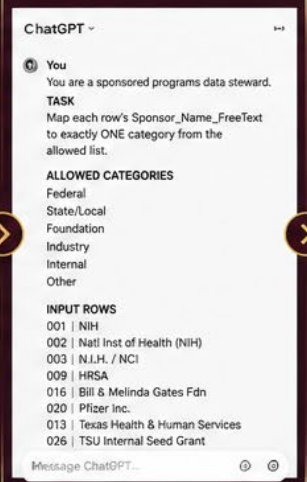
Example of inconsistent sponsor-name entries.

2 Copy Data (From Excel)

Row_ID	Sponsor_Name_FreeText
001	NIH
002	Natl Inst of Health (NIH)
003	N.I.H. / NCI
009	HRSA
016	Bill & Melinda Gates Fdn
020	Pfizer Inc.
013	Texas Health & Human Services
026	TSU Internal Seed Grant

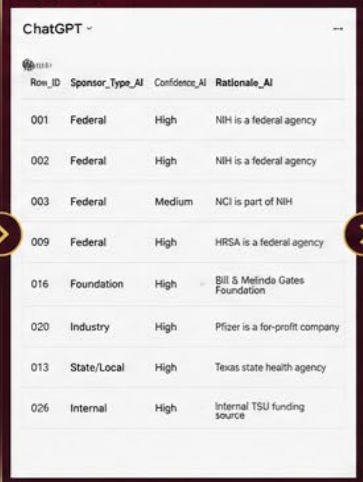
Copy the 8 rows to use in the prompt.

3 Paste into ChatGPT (With Prompt)



Paste the prompt with the 8 rows and submit.

4 AI Returns Structured Suggestions



Row_ID	Sponsor_Type_AI	Confidence_AI	Rationale_AI
001	Federal	High	NIH is a federal agency
002	Federal	High	NIH is a federal agency
003	Federal	Medium	NCI is part of NIH
009	Federal	High	HRSA is a federal agency
016	Foundation	High	Bill & Melinda Gates Foundation
020	Industry	High	Pfizer is a for-profit company
013	State/Local	High	Texas state health agency
026	Internal	High	Internal TSU funding source

The model provides category, confidence, and rationale.

5 Paste Back to Excel & Review

Row_ID	Sponsor_Name_FreeText	Sponsor_Type_AI	Confidence_AI
001	NIH	Federal	High
002	Natl Inst of Health (NIH)	Federal	High
003	N.I.H. / NCI	Federal	Medium
009	HRSA	Federal	High
016	Bill & Melinda Gates Fdn	Foundation	High
020	Pfizer Inc.	Industry	High
013	Texas Health & Human Services	State/Local	High
026	TSU Internal Seed Grant	Internal	High

Paste the results and check the confidence.

6 Human Review & Mapping Dictionary

Row_ID	Human_Decision	Reviewer	Decision_Date	Notes
001	Accept	JSMITH	5/12/2025	
002	Accept	JSMITH	5/12/2025	
003	Accept	JSMITH	5/12/2025	
009	Accept	JSMITH	5/12/2025	
016	Accept	JSMITH	5/12/2025	
020	Accept	JSMITH	5/12/2025	
013	Accept	JSMITH	5/12/2025	
026	Accept	JSMITH	5/12/2025	

Mapping Dictionary (Verified)

Sponsor_Name_FreeText	Sponsor_Type	Verified_By	Verified_Date
NIH	Federal	JSMITH	5/12/2025
Natl Inst of Health (NIH)	Federal	JSMITH	5/12/2025
N.I.H. / NCI	Federal	JSMITH	5/12/2025
HRSA	Federal	JSMITH	5/12/2025
Bill & Melinda Gates Fdn	Foundation	JSMITH	5/12/2025
Pfizer Inc.	Industry	JSMITH	5/12/2025
Texas Health & Human Services	State/Local	JSMITH	5/12/2025
TSU Internal Seed Grant	Internal	JSMITH	5/12/2025

Reviewer validates. Verified patterns are saved for reuse.

GOVERNED

Controlled categories
Human oversight
Audit-ready

EFFICIENT

Reduces repetitive work
Improves consistency
Reusable mappings

SUSTAINABLE

Ruilds institutional memory
Strengthens data quality
Supports better reporting

Simple workflow.
Strong governance.
Trusted outcomes.

KEY TAKEAWAYS

NEAT
No-code Entry Alignment Toolkit

THE VALUE IS NOT AUTOMATION ALONE.

- Preserving judgment.
- Reducing repeated cleanup.
- Capturing institutional knowledge.
- Creating more time for higher-value work.

SELECTED REFERENCES

- Cayuse. (2025a). *Conducting accurate workload assessments in research administration*. Cayuse.
- Cayuse. (2025b). *7 challenges with homegrown research administration solutions*. Cayuse.
- National Science and Technology Council. (2024). *Advancing research capacity at high research activity Historically Black Colleges and Universities*. Executive Office of the President.
- Woodward, J. E., & Roberts, E. (2025). *The needs and challenges of the research administrator workforce*. Journal of Research Administration, 56(2), 7–19.

SCAN THE QR CODES TO CONNECT ON LINKEDIN

Thank You

Mana Rassaeikashuk, MSc, MBA
Doctoral Student | Pharmaceutical Sciences @ Mana Kashuk

Dr. Esther Olaleye, PhD, MPH
Senior Associate Vice President for Research and Innovation | Professor of Pharmacology and Epidemiology @ Esther Olaleye

TSU Texas Southern University Division of Research & Innovation | Texas Southern University Joan M. Lafleur College of Pharmacy & Health Sciences | CBMHR Center for Biomedical & Minority Health Research Community Engagement Core