

A Three-GenAI Consensus Model for Deductive Analysis of Response Data

Amruth N. Kumar, Renee McCauley,
Bonnie K. MacKellar, Tammy VanDeGrift

Qualitative Research

- Inductive Coding
 - Text data → Codes → Themes → Codebook
- Deductive Coding
 - Text Data + Codebook → Theme associated with each data item

Deductive Coding

- Multiple researchers
- Trained
- Multiple Iterations
- Consensus
- Expensive
 - *Time*
 - *Effort*
- Does not scale well

Three-GenAI Consensus Model

- Split shuffled data into Trial (10-15%) and Corpus (85-90%)
- Draft a Codebook-based prompt
- Iteratively revise prompt:
 - Code trial data
 - Compare against ground truth (IRR)
 - Revise prompt
- Have 3 GenAIs code corpus data using revised prompt

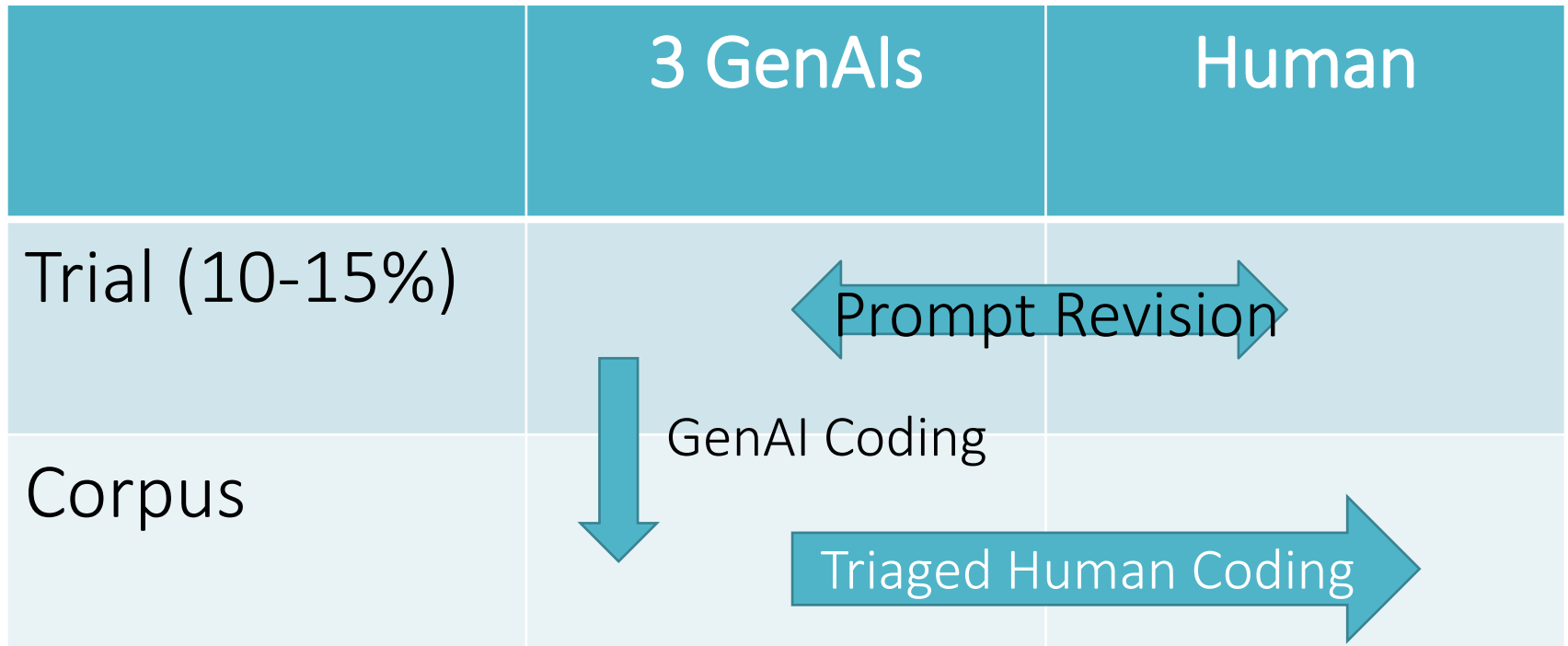
Tabulate GenAI Responses

- Unanimous
- One category per coding unit
- 2-Agree
- Two categories per coding unit
- 3-Disagree
- Three categories per coding unit

Deductive coding of corpus data

- Consider the 1 / 2 / 3 categories first
- If correct, assign
- If incorrect, consider the remaining categories in the codebook

Process Summarized



Dispositions

- Competency = Knowledge + Skills + Dispositions
- Dispositions: Willingness and intent to apply knowledge and skills to solve problems
- Dispositions important for computer science [CS2023]
- **Self-directed**
- Persistent
- Meticulous
- Collaborative

Self-Directed

Definition: Learn new tools/techniques on your own to complete a task, even when the tool/technique is only minimally used or discussed in class, and you may not receive additional credit just for learning it

Prompt: Describe an example of you being self-directed when completing this assignment. Otherwise, describe the circumstances that prevented you from being self-directed.

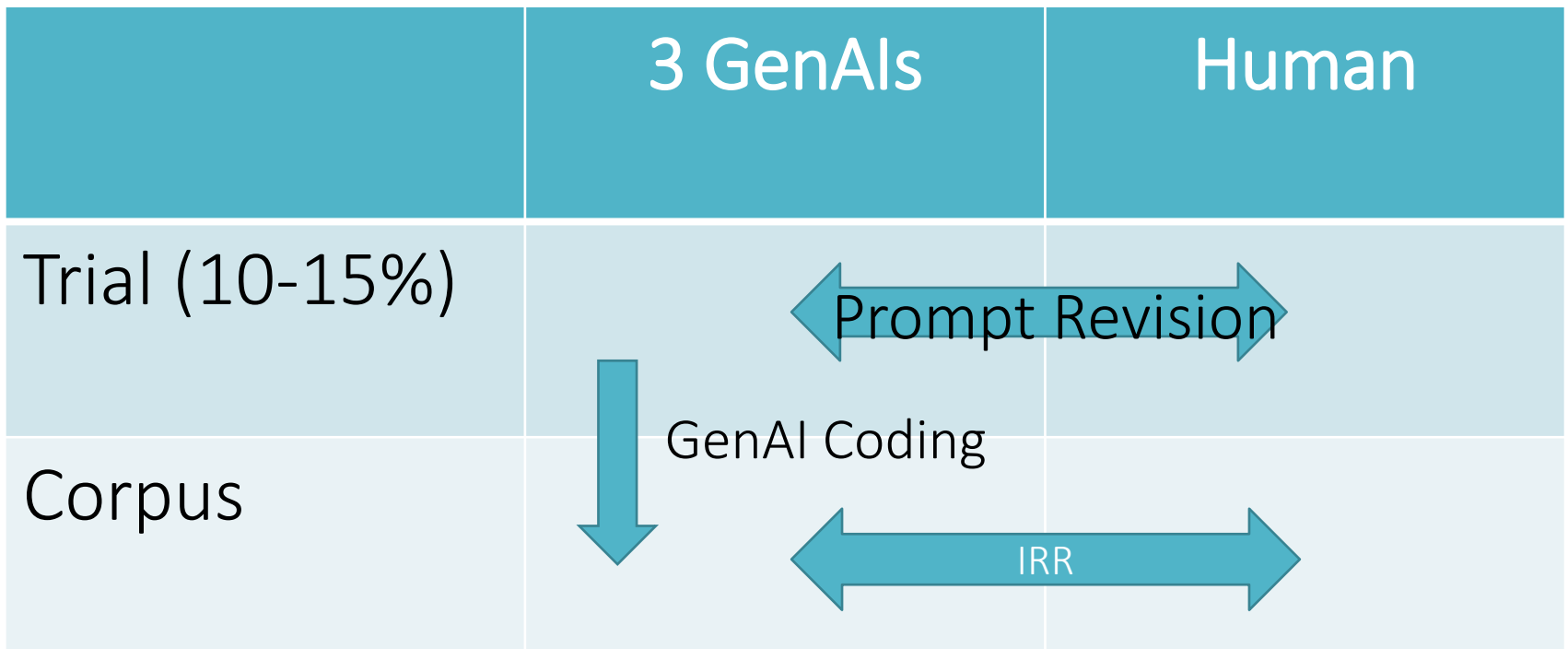
Self-Directed data

- 2022-2024
- 5 computing courses
- 4 institutions
- 264 consenting students
- 496 responses
- Prior coding = Ground truth (for proof-of-concept)
- By human researchers

Codebook Categories

- Utilizing external resources
- Learning necessary material
- Working independently
- Assessing oneself
- Planning ahead
- Applying useful techniques
- Completing the assigned work
- Reviewing against expectations
- Not applied / Not pertinent

Case Study Summarized



Unanimous Cluster

Data	Total N	Unanimous N	Ground Truth N	Ground Truth %
Trial	74	52	42	81% 0.75
Corpus	422	289	238	82% 0.77
Total	496	341	280	82%

2-Agree Cluster

Data	Total N	2-Agree N	Ground Truth N	Ground Truth %
Trial	74	18	6 + 5	61%
Corpus	422	109	57 + 32	82%
Total	496	127	100	79%

3-Disagree Cluster

Data	Total N	3-Disagree N	Ground Truth N	Ground Truth %
Trial	74	4	1	25%
Corpus	422	24	20	83%
Total	496	28	21	75%

Time Savings

Percentage reduction in time:

$$n_u * p_u * (c - 1) + n_a * p_a * (c - 2) + n_d * p_d * (c - 3)$$

$$S * C$$

Time Savings in Case Study

- $n_u = 341, n_a = 127, n_d = 28, s = 496$

- $p_u = 0.82, p_a = 0.79, p_d = 0.75$

- $c = 10$



69.88% (minus time spent to use the model)

Summary

- Consensus used, not individual response
- Consensus used to triage data
- Researchers still make the final call for *every* data item



Reallocate unanimous / 2-agree time to 3-Disagree data

Future Work

- Formalize human review

 - Human arbitration

- Apply to additional data

 - Persistent

 - Meticulous

- Zero-shot → One-shot prompt

 - Leave-one-out approach

Acknowledgments



Partial support provided by
National Science Foundation under grants
DUE-2216121, 2142648

dispositions-project.org

Interest welcome!