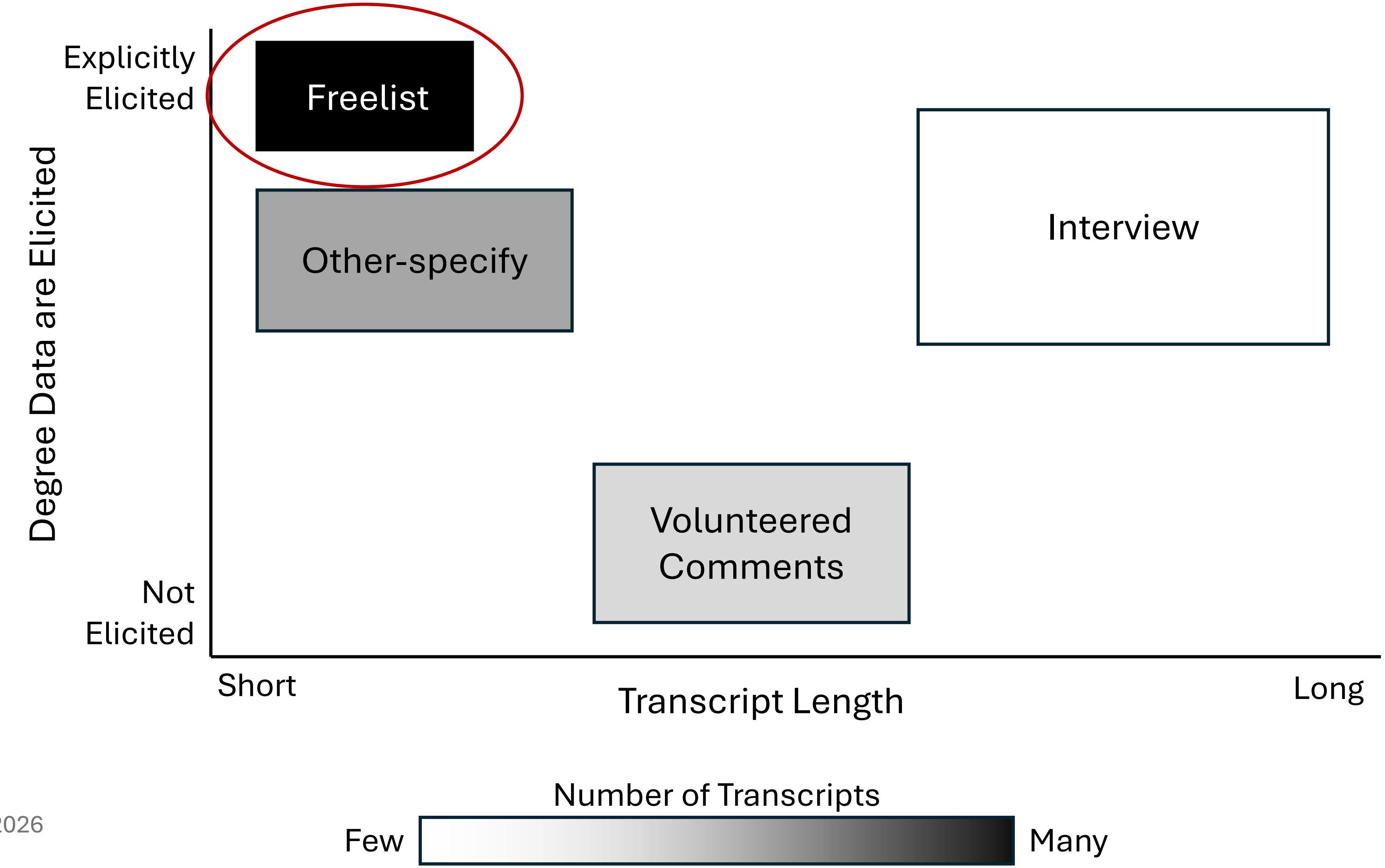
Coding at Scale: Human-LLM Partnerships in Large-Sample Qualitative Research

Andrew Parker, Kelly Hyde, Gary Mottola, Olivia Valdes, Duygu Basaran Sahin, Emily Hoch

Funded by the FINRA Investor Education Foundation

Presented at CIPHER 2026

Motivation: Scalable Qualitative Coding (AI)



Use Case: Freelisting on Large-sample Surveys

[2015] When you think of Zika, what other diseases come to mind? Please list up to three other diseases.

1.
2.
3.

[2024] What could a person do to protect themselves from fraud? Please list up to three behaviors in the text boxes below (one per box):

1.
2.
3.

RAND American Life Panel

Survey 608

1509 respondents

(58.7% completion rate)

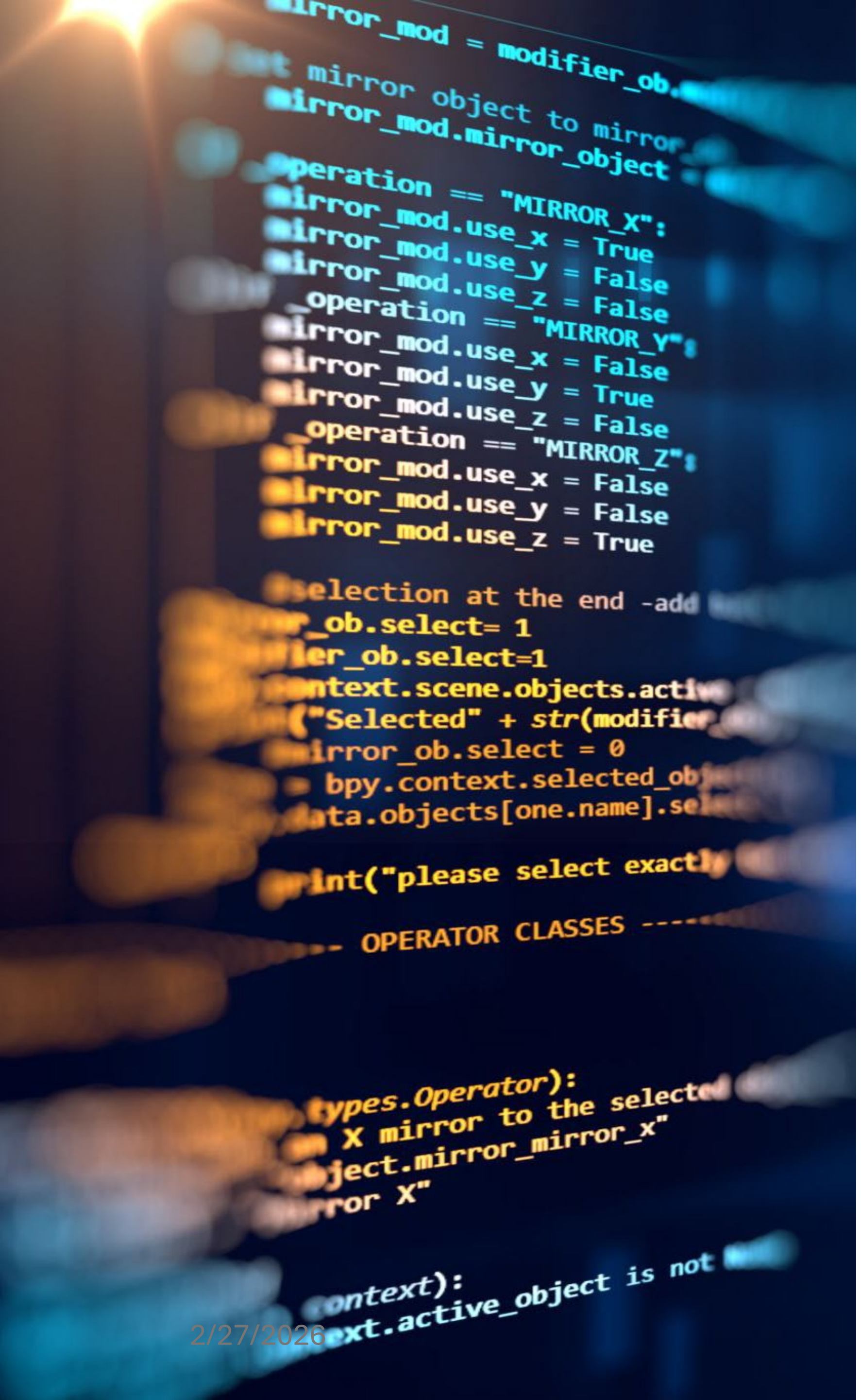
3810 total entries

(avg 2.5 responses per respondent)

6.7 words and 39.3 characters, on average

Embedded within larger ALP survey

Linked to all other ALP surveys



"MuseQDA as a tireless, naïve, and abused graduate student"

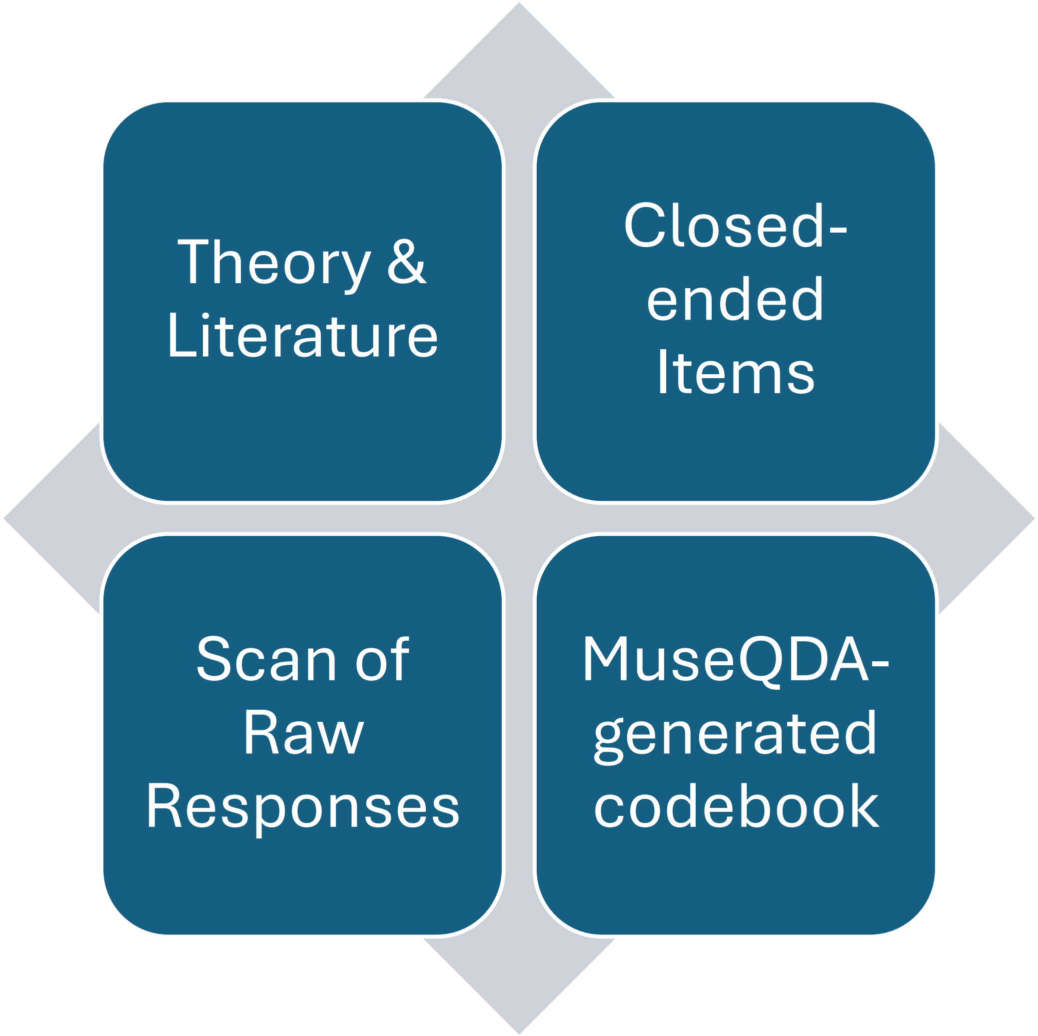
1. Humans pose initial codebook
2. Multiple human coders review a sample of transcripts, revise codebook as appropriate
3. Upload codebook into MuseQDA and have Muse code all transcripts
4. Humans review coding and adjust codebook, reapplying codes in MuseQDA
5. Human coders apply updated codebook to new random set of entries, checking inter-rater reliability
6. Rinse and repeat as necessary
7. Statistically compare resulting codes to other variables of interest

Construct Validation

- Domain of Observables
- Relations among Observables
- Relations among Constructs

Nunnally, J.C., & Bernstein, I.H. (2007). *Psychometric Theory*. McGraw-Hill.

Domain of Observables:
Codebook generation draws on multiple sources



Relations among Observables: Inter-Rater Reliability … and Intra?

Code	Human-Human Kappa	Human-Muse Kappa	Human Base Rate	Machine Base Rate
Don't Engage	0.69	**0.80**	29%	31%
Don't Share Personal Information	0.57	**0.61**	23%	19%
Skepticism, Awareness, and Seeking Information	0.68	**0.73**	13%	20%
Verify or Check Identities	0.48	**0.76**	13%	10%
Account and Device Security	**0.83**	0.67	9%	16%
Fraud Protection Tools	**0.76**	0.61	8%	9%
Monitor Accounts and Credit Reports	0.76	**0.92**	3%	3%
Reporting Fraud	0.75	**0.89**	2%	2%

Note: n=202; Bold indicates higher kappa

Relations among Constructs
Predicting protective behavior with fraud type

	Don't Engage	Don't Share Personal Information	Skepticism, Awareness, and Seeking Information	Verify or Check Identities	Account and Device Security	Fraud Protection Tools	Monitor Accounts and Credit Reports	Reporting Fraud
Opportunity-based Fraud	1.038	0.920	**2.266*** **	1.459	0.909	0.890	**0.311***	1.764
	(0.283)	(0.228)	(0.544)	(0.367)	(0.221)	(0.263)	(0.112)	(0.661)
Consumer-based Fraud	1.156	1.034	1.310	1.463	0.678	**0.526****	1.180	1.145
	(0.221)	(0.210)	(0.279)	(0.308)	(0.155)	(0.143)	(0.351)	(0.392)
ID-based Fraud	1.206	**1.688***	0.807	1.288	**3.036*****	**3.621*****	**2.859***	1.328
	(0.267)	(0.337)	(0.157)	(0.283)	(0.690)	(0.937)	(1.001)	(0.616)
Imposter-based Fraud	1.153	**1.799****	1.524	**2.625*****	0.906	0.635	**0.388****	0.576
	(0.297)	(0.456)	(0.365)	(0.667)	(0.251)	(0.191)	(0.161)	(0.169)
Threat-based Fraud	1.326	1.750	0.761	1.441	0.981	1.386	1.310	0.982
	(0.456)	(0.514)	(0.209)	(0.567)	(0.381)	(0.579)	(0.590)	(0.312)

Note: N=1509; odds ratios from logistic regressions predicting protective behaviors and controlling for demographics.
* two-sided p-value < .05; ** p < .01; *** p < .001

Next Steps

- Coding complex stories as separate parent-child branches
  - Fraud mode
  - Fraud type
  - Behavioral responses
  - Emotional reactions
- Near real-time incorporation of open-ended public perspectives
  - Longitudinal assessment