



Leveraging Artificial Intelligence to Code “Other/Specify” Responses in Surveys

CIPHER 2026

2/27/2026

Ting Yan, Brandon Sepulvado, Lilian Huang, Roopam Singh, Erlina Hendarwan, Enoch Platero, Ipek Bilgen, David Dutwin



Three methods to code open-ended answers, each with pros and cons

Human Coding

Most accurate and reliable; but time consuming and labor intensive

Machine Learning (ML) methods

Automated to reduce cost and time; but time and cost for preprocessing, training, and fine-tuning

Large Language Models (LLMs)

Potential to further reduce cost and time; but quality depends on models and prompts

- Respondents are offered a list of response options to choose from
- Respondents are also offered a textbox
 - to enter additional information, or
 - to specify answer that do not fit predefined response options
- But verbatim answers tend to be very brief

7. What is this person's race?Mark one or more boxes **AND** print origins.

- White – Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc. ↴

- Black or African Am. – Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc. ↴

- American Indian or Alaska Native – Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc. ↴

- | | | |
|---|-------------------------------------|--|
| <input type="checkbox"/> Chinese | <input type="checkbox"/> Vietnamese | <input type="checkbox"/> Native Hawaiian |
| <input type="checkbox"/> Filipino | <input type="checkbox"/> Korean | <input type="checkbox"/> Samoan |
| <input type="checkbox"/> Asian Indian | <input type="checkbox"/> Japanese | <input type="checkbox"/> Chamorro |
| <input type="checkbox"/> Other Asian –
Print, for example,
Pakistani, Cambodian,
Hmong, etc. ↴ | | <input type="checkbox"/> Other Pacific Islander –
Print, for example,
Tongan, Fijian,
Marshallese, etc. ↴ |

- Some other race – Print race or origin. ↴

Please check one or more categories below to indicate what race or races you consider yourself to be. Select all that apply.

01. White

02. Black or African American

03. American Indian or Alaska Native – Type in name of enrolled or principal tribe

TEXTBOX

04. Asian Indian

05. Chinese

06. Filipino

07. Japanese

08. Korean

09. Vietnamese

10. Other Asian – for example, Bangladeshi, Burmese, Cambodian, Hmong, Indonesian, Laotian, Malaysian, Pakistani, Sri Lankan, Taiwanese, Thai, etc. Type in race **TEXTBOX**

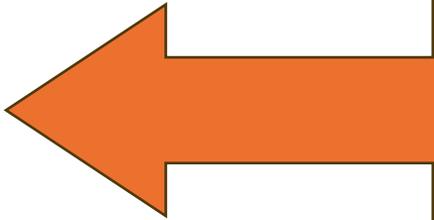
11. Native Hawaiian

12. Guamanian or Chamorro

13. Samoan

14. Other Pacific Islander – for example, Fijian, Tongan, etc. Type in race **TEXTBOX**

15. Some other race – Type in race **TEXTBOX**

- 
- 1) Data are captured as 15 variables with 0/1 value
 - 2) Answers to other/specify options need to be coded into one of the response options
 - 3) Coded values may affect the values of other 0/1 variables

Please check one or more categories below to indicate what race or races you consider yourself to be. Select all that apply

01. White

02. Black or African American

03. American Indian or Alaska Native – Type in name of enrolled or principal tribe

TEXTBOX

04. Asian Indian

05. Chinese

06. Filipino

07. Japanese

08. Korean

09. Vietnamese

10. Other Asian – for example, Bangladeshi, Burmese, Cambodian, Hmong, Indonesian, Laotian, Malaysian, Pakistani, Sri Lankan, Taiwanese, Thai, etc. Type in race **TEXTBOX**

11. Native Hawaiian

12. Guamanian or Chamorro

13. Samoan

14. Other Pacific Islander – for example, Fijian, Tongan, etc. Type in race **TEXTBOX**

15. Some other race – Type in race **TEXTBOX**

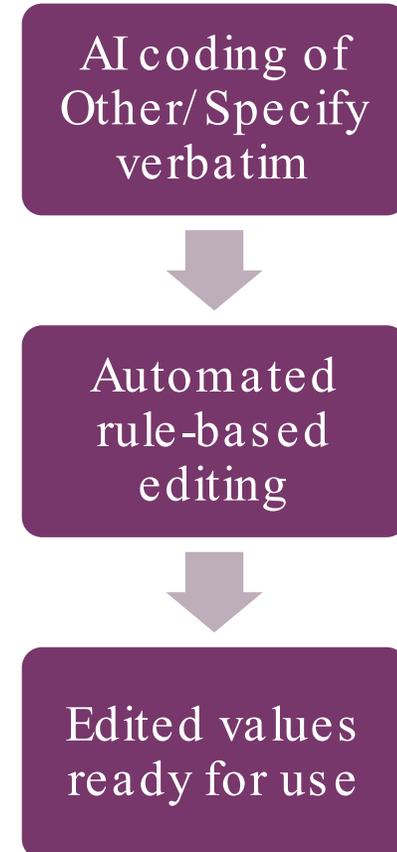
- R entered “Taiwanese” to other/specify box to “Other Asian”
- Race_05 need to be coded to 1 and Race_10 needed to be coded to 0
- Human coding tends to be case level

Please check one or more categories below to indicate what race or races you consider yourself to be. Select all that apply

01. White
02. Black or African American
03. American Indian or Alaska Native – Type in name of enrolled or principal tribe
TEXTBOX
04. Asian Indian
05. Chinese
06. Filipino
07. Japanese
08. Korean
09. Vietnamese
10. Other Asian – for example, Bangladeshi, Burmese, Cambodian, Hmong, Indonesian, Laotian, Malaysian, Pakistani, Sri Lankan, Taiwanese, Thai, etc. Type in race **TEXTBOX**
11. Native Hawaiian
12. Guamanian or Chamorro
13. Samoan
14. Other Pacific Islander – for example, Fijian, Tongan, etc. Type in race **TEXTBOX**
15. Some other race – Type in race **TEXTBOX**

Can LLMs be leveraged to increase efficiency of “other/specify” race coding?

- A workflow separating the coding and editing tasks into phases
 - Phase 1 uses AI to code other/specify verbatims into one of the 15 options and stores the coded values
 - Phase 2 determines whether the values of other variables need to be updated or not based on a set of business rules
 - Phase 3 output edited values for evaluation and derivation of new variables
- This workflow codes/edits values at the variable level (instead of case level)
- This setup makes it easy to identify where problems happen and easy to update specific procedures (e.g., using a different prompt at Phase 1 or a new set of business rules at Phase 2)

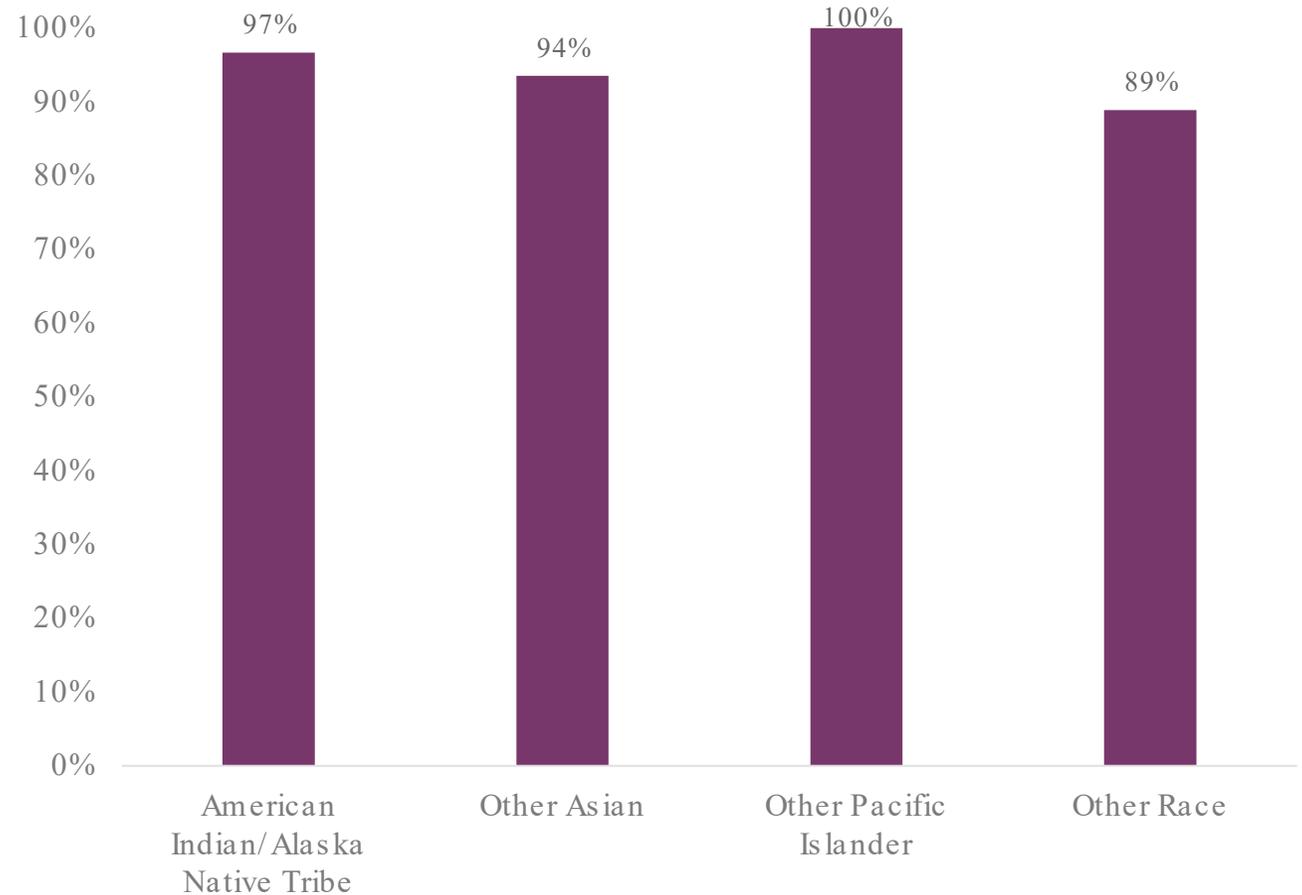


- Selected 100 panelists who provided at least one other/specify answer to the race question
- The other/specify answers have been coded by staff and race variables have been edited by staff
- Sent other/specify answers for AI coding
 - Zero-shot prompting: code frame + response text
 - Model: GPT-4o with temperature = 0
- Coded values were used for editing of other variables using Python
- The AI coded values were reviewed by three senior staff
- The edited values output from the workflow were compared to values edited by staff
- Disagreements were reviewed by three senior staff

- AI did a great job coding other/specify answers to one of the race options

- Mistakes

- “Mayan” entered to the AIAN tribe option coded as “other race”
- “Taiwanese American” entered to the “Other Asian” coded as “Other Asian”
- “Afghan” entered to the “Other Asian” coded “Middle Eastern and North African”
- “98” entered to “Other Race” and AIAN tribe coded as “Uncodeable”

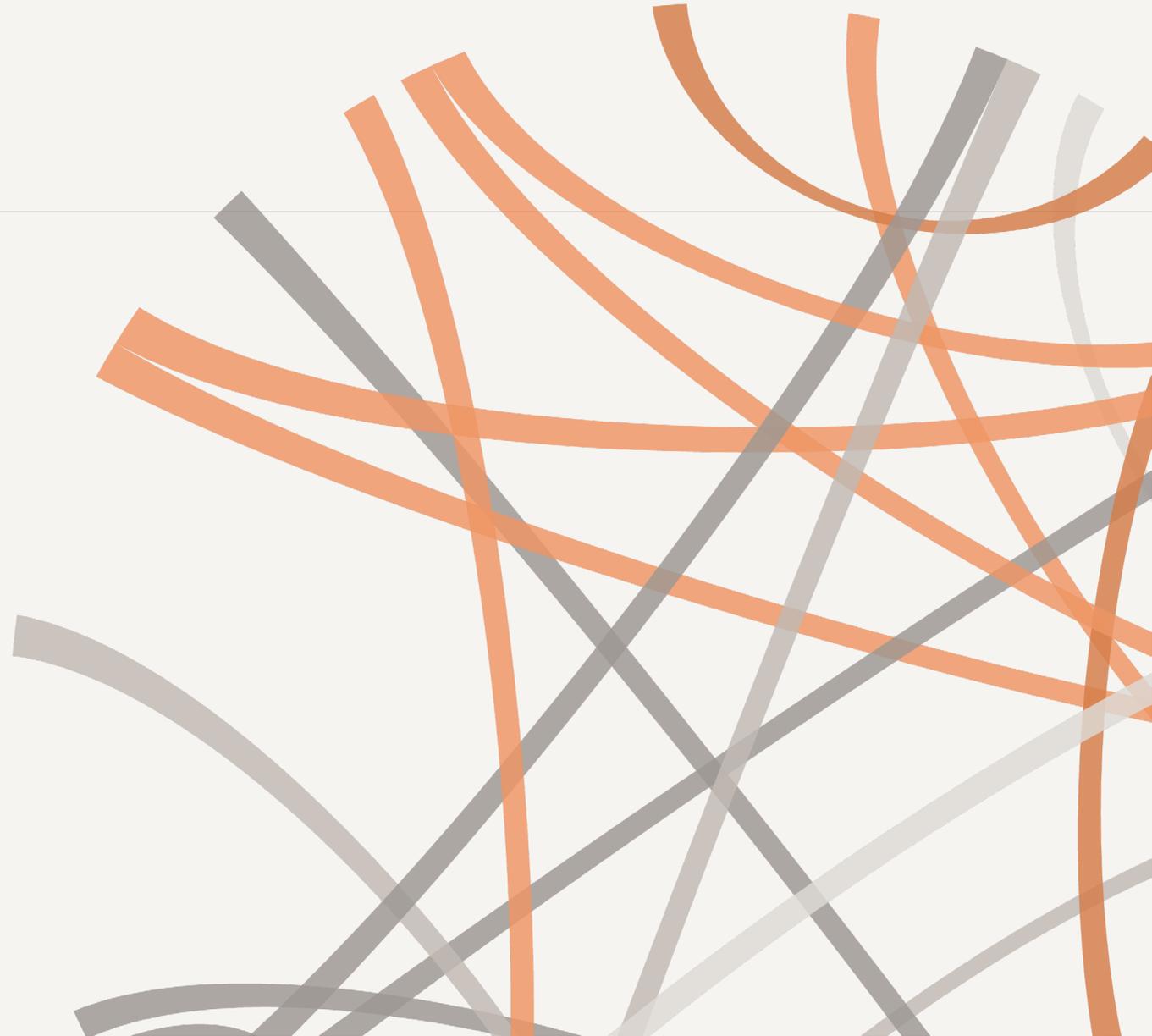


- The workflow implemented business rules more consistently than staff
 - Especially for variables associated with other/specify options
- The sequential flow also allowed us to reconsider some business rules and to easily added new flags for “Middle Eastern and Northern African”

Key Takeaways and Next Steps

- The AI-enabled workflow has a great potential to improve efficiency and consistency
 - AI coding takes .7 sec/response and automated editing routine takes < 1 minute to run
- AmeriSpeak is upscaling the AI-enabled workflow
- AmeriSpeak is exploring to improve the accuracy of AI coding
 - Asking LLMs to output “certainty”/ “confidence” measure of their own coding
 - Considering to provide examples to clarify ambiguities (i.e., few-shot prompting)
 - Finetuning small language models (SLMs)
 - Having multiple LLMs or SLMs “vote” on the most appropriate code/category
 - Also working on using LLMs to clean/preprocess raw text

Questions?



Thank you.

Ting Yan
Yan-ting@norc.org

Get Your Research Right

