Appendix

A   Prototype JSONL entry for the Batch API request for scoring

{"custom_id": "ABC123XYZ890", "method": "POST", "url":
"/v1/chat/completions", "
body": {"model": "gpt-4o", "n": 1, "max_tokens": 1, "presence_penalty": 0,
"frequency_penalty": 0, "temperature": 0.000001, "top_p": 0.1, "messages":
[{"role": "system", "content": "# Holistic Rating Form

After reading each essay and completing the analytical rating form, assign a holistic score based on the rubric below. For the following evaluations you will need to use a grading scale between 1 (minimum) and 6 (maximum). As with the analytical rating form, the distance between each grade (e.g., 1-2, 3-4, 4-5) should be considered equal.

SCORE OF 6: An essay in this category demonstrates clear and consistent mastery, although it may have a few minor errors. A typical essay effectively and insightfully develops a point of view on the issue and demonstrates outstanding critical thinking, using clearly appropriate examples, reasons, and other evidence to support its position; the essay is well organized and clearly focused, demonstrating clear coherence and smooth progression of ideas; the essay exhibits skillful use of language, using a varied, accurate, and apt vocabulary and demonstrates meaningful variety in sentence structure; the essay is free of most errors in grammar, usage, and mechanics.

SCORE OF 5: An essay in this category demonstrates reasonably consistent mastery, although it will have occasional errors or lapses in quality. A typical essay effectively develops a point of view on the issue and demonstrates strong critical thinking, generally using appropriate examples, reasons, and other evidence to support its position; the essay is well organized and focused, demonstrating coherence and progression of ideas; the essay exhibits facility in the use of language, using appropriate vocabulary demonstrates variety in sentence structure; the essay is generally free of most errors in grammar, usage, and mechanics.

SCORE OF 4: An essay in this category demonstrates adequate mastery, although it will have lapses in quality. A typical essay develops a point of view on the issue and demonstrates competent critical thinking, using adequate examples, reasons, and other evidence to support its position; the essay is generally organized and focused, demonstrating some coherence and progression of ideas exhibits adequate; the essay may demonstrate inconsistent facility in the use of language, using generally appropriate vocabulary demonstrates some variety in sentence

structure; the essay may have some errors in grammar, usage, and mechanics.

SCORE OF 3: An essay in this category demonstrates developing mastery, and is marked by ONE OR MORE of the following weaknesses: develops a point of view on the issue, demonstrating some critical thinking, but may do so inconsistently or use inadequate examples, reasons, or other evidence to support its position; the essay is limited in its organization or focus, or may demonstrate some lapses in coherence or progression of ideas displays; the essay may demonstrate facility in the use of language, but sometimes uses weak vocabulary or inappropriate word choice and/or lacks variety or demonstrates problems in sentence structure; the essay may contain an accumulation of errors in grammar, usage, and mechanics.

SCORE OF 2: An essay in this category demonstrates little mastery, and is flawed by ONE OR MORE of the following weaknesses: develops a point of view on the issue that is vague or seriously limited, and demonstrates weak critical thinking, providing inappropriate or insufficient examples, reasons, or other evidence to support its position; the essay is poorly organized and/or focused, or demonstrates serious problems with coherence or progression of ideas; the essay displays very little facility in the use of language, using very limited vocabulary or incorrect word choice and/or demonstrates frequent problems in sentence structure; the essay contains errors in grammar, usage, and mechanics so serious that meaning is somewhat obscured.

SCORE OF 1: An essay in this category demonstrates very little or no mastery, and is severely flawed by ONE OR MORE of the following weaknesses: develops no viable point of view on the issue, or provides little or no evidence to support its position; the essay is disorganized or unfocused, resulting in a disjointed or incoherent essay; the essay displays fundamental errors in vocabulary and/or demonstrates severe flaws in sentence structure; the essay contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning.
"},{"role": "user", "content": "score the essay that follows using the holistic scoring rubric: give only the numeric score. do not write anything other than the number. The essay is:
<STUDENT ESSAY GOES HERE!>}]}}

B   Prototype JSONL entry for the Batch API request for race/ethnicity prediction

{"custom_id": "ABC123XYZ890", "method": "POST", "url": "/v1/chat/completions", "body": {"model": "gpt-4o", "n": 1, "max_tokens": 1, "presence_penalty": 0, "fr equency_penalty": 0, "temperature": 1e-06, "top_p": 0.1, "messages": [{"role": " system", "content": "The race ethnicity of the writer of the essay below is one

of : (1) American Indian/Alaskan Native; (2) Asian/Pacific Islander;
(3) Black/A frican American; (4) Hispanic/Latino; (5) Two or more
races/Other; (6) White. I want you to predict the race ethnicity of
the author. The essay is:

<STUDENT ESSAY GOES HERE!>"},
{"role": "user", "content": "Provide
your prediction of the race ethnicity using the numeric codes above.
Do not write any words in your response to me."}]}}