def create\_chatcompletion\_messages(row):

return [

{

"role": "system",

"content": "As a Biological Data Analyst with expertise in demographic categorization and human populations, your primary task is to standardize the 'Race' field. You have good knowledge in human development and social classifications, particularly when referring to various racial and ethnic groups."

},

{

"role": "user",

"content": f"Carefully review the info below: {row['Race']}\n"

"Before starting the standardization, check the total number of Original Terms provided. Ensure that the order of terms in your Standardized terms matches the exact order in the input data. Standardize each Original Term sequentially without omission, and verify that the total number of Standardized Terms matches the count in the input.\n"

"Detailed Instructions for Improving Revised Terms:\n"

"1. Maintain Consistency: Always use commonly accepted terms for race and ethnicity. Ensure that terms are clear, concise, and widely recognized in demographic and scientific contexts.\n"

"2. No Abbreviations: Avoid abbreviations unless they are universally understood, such as 'NA' for 'Not Available'. 'AI/AN' should be expanded to 'American Indian or Alaska Native.'\n"

"3. Avoid Using Multiple Terms: Where there is a combination of racial categories (e.g., 'Black or African American; White'), standardize to a single category or use 'Mixed' or 'Multi-racial' when necessary.\n"

"4. Keep 'NA' as it is: Do not expand 'NA' to 'Not Applicable' or any other variants. Always retain 'NA' as the value when it appears.\n"

"5. Simplify Complex Terms: For terms like 'White; American Indian or Alaska Native,' standardize to a clear format like 'White + American Indian or Alaska Native.'\n"

"6. Use Recognized Terms for Mixed Groups: For terms indicating more than one race, standardize them as 'Mixed' or 'Multi-racial' as needed, and avoid ambiguous descriptions like 'More than one race.'\n"

"7. Gender Neutrality: Ensure that terms are neutral and inclusive. For example, terms like 'Hispanic' and 'Latino' should be standardized to 'Hispanic or Latino' to be inclusive.\n"

"8. Geographic Clarity: Ensure that terms referring to geographic regions, such as 'South Indian' or 'Middle Eastern,' are standardized to be more broadly accepted, e.g., 'Indian' and 'Middle Eastern' respectively.\n"

"9. Ethnicity Clarification: Ensure that ethnic groups with unique identifiers, such as 'Native Hawaiian or Pacific Islander,' are standardized and unambiguous.\n"

"10. Don't Use 'Not Specified': Replace 'Not Specified' with 'NA' to indicate missing or unknown data.\n"

"11. Hispanic/Latino Clarification: Use 'Hispanic or Latino' for any variation of these terms to maintain consistency.\n"

"12. Clarify Unknown Data: Replace 'Unknown/Not reported' with 'Unknown' for missing or unspecified data.\n"

"Format the Output as a Table:\n"

" - Create a table with two columns:\n"

" - List each original term separately in the first column.\n"

" - Provide the standardized term in the second column, ensuring clarity and consistency.\n"

" - Apply domain expertise to ensure accuracy in the standardization process.\n"

"Examples:\n"

"| Original Term | Standardized Term |\n"

"|-----------------------------------------|-------------------------------------------|\n"

"| African | African |\n"

"| African American | African American |\n"

"| AI/AN | American Indian or Alaska Native |\n"

"| American Indian or Alaska Native | American Indian or Alaska Native |\n"

"| American Indian or Alaska Native; White | American Indian or Alaska Native + White |\n"

"| Black | Black |\n"

"| Black or African American | Black or African American |\n"

"| Hispanic | Hispanic or Latino |\n"

"| Hispanic or Latino | Hispanic or Latino |\n"

"| Middle Eastern | Middle Eastern |\n"

"| Mixed | Mixed |\n"

"| Multi-racial | Multi-racial |\n"

"| Native American | Native American |\n"

"| Native Hawaiian or Pacific Islander | Native Hawaiian or Pacific Islander |\n"

"| NA | NA |\n"

"| Other | Other |\n"

"| White | White |\n"

"| White; Asian | White + Asian |\n"

"| White; Other | White + Other |\n"

"| White/Asian | White or Asian |"

}

]