def create\_chatcompletion\_messages(row):

return [

{

"role": "system",

"content": "You are a Bioinformatics Annotator and Treatment Standardization Specialist. Your task is to standardize the treatment dose field based on the provided dataset. The primary objective is to ensure accurate identification, standardization, and consistency of treatment doses, taking into account capitalization, unit distinctions, and proper formatting."},

{

"role": "user",

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

"Guidelines for Standardization of the 'Treatment\_Dose' Field:\n"

"1. Units of Measurement: Ensure that all terms use consistent units.\n"

" - Use 'µg' (micrograms) instead of 'ug' to reflect standard scientific notation.\n"

" - Use 'µM' (micromolar) instead of 'uM.'\n"

" - Maintain capitalization for proper units like 'nM' for nanomolar and 'mM' for millimolar.\n"

" - Convert all other unit terms to standard scientific abbreviations (e.g., 'ml' for milliliters).\n"

" - Replace 'kcal' for calorie-related terms. For example, '60% calories' → '60% kcal'.\n"

" - Standardize 'genomic copies' to 'genome copies' (e.g., '2 × 10^11 genomic copies/mouse' → '2 × 10^11 genome copies/mouse').\n\n"

" - For any mentions of 'body weight' or 'Body Weight', use 'bw' consistently (e.g., 'µg/g body weight/day' → 'µg/g bw/day').\n\n"

"2. Special Units: Ensure specific units like CFU and PFU are used consistently without changes. Do not add an 's' at the end, even in plural contexts."

" - Colony Forming Units (CFU): Maintain the abbreviation 'CFU.' If additional clarification is required (e.g., bacterial species or context), provide it in the comment field.\n"

" - Plaque Forming Units (PFU): Maintain the abbreviation 'PFU.' Clarify viral species or treatment context if necessary.\n"

" - Replace '1 x 10^8 cfu' or similar variations with '1 x 10^8 CFU' for capitalization consistency.\n"

" - Use 'MOI' for multiplicity of infection, ensuring consistency. For example, 'moi 2' → 'MOI 2'; '3 MOI' → 'MOI 3'.\n"

" - Use 'focus forming units' for similar contexts. For example, '2 x 10^6 focal-forming units' → '2 x 10^6 focus forming units'.\n"

"MOI: Replace '=' with a space."

"Example: 'MOI=1' → 'MOI 1'"

"Percent Volume/Volume (v/v):"

"Remove parentheses but retain 'v/v' format."

"Example: '1% (v/v)' → '1% v/v'"

"3. Scientific Notation and Spacing: Standardize scientific notation across the dataset.\n"

" - Convert 'e' notation to standard power notation (e.g., '10e5 cells' becomes '10^5 cells').\n"

" - Use lowercase 'x' with spacing for scientific notation, e.g., '1x10^6' → '1 x 10^6'.\n"

" - Maintain consistent formatting for exponents (e.g., '1 x 10^5 cells').\n\n"

"4.Preserve Order and Structure: The order in which the dosages appear should be preserved as they represent a sequence of treatment. The format should be: Dosage Value + Unit, separated by + when there are multiple entries."

"Multiple Dosages and Combinations: For terms with multiple doses or combinations, separate them using ' + '.\n"

"Do not split terms connected by a semicolon. If a semicolon is present, it indicates that the term is complex and should remain unified as one single term in subsequent steps for clarity and consistency."

"Ensure spacing around units for readability (e.g., '100 µg/ml + 10 ng/ml').\n"

"Example: 5000 U/ml + 100 U/ml + 10 µg/ml + 10 µg/ml."

"5. Consistency in Capitalization: Ensure that units and measurements are capitalized correctly.\n"

" - For example, '100ng/ml' should be converted to '100 ng/ml'.\n"

" - '10e5 cells' should be converted to '10^5 cells.'\n"

" - '100ug/ml' should be standardized as '100 µg/ml.'\n"

" - '100uM' should be converted to '100 µM.'\n"

" - '10^7 CFU' should be retained unless additional context is required.\n"

"6. Remove Drug Names: Exclude any drug or chemical names preceding the dosage. Only the numerical values with their corresponding units (such as U/ml, ng/ml, µg/ml, etc.) should remain in the annotation. Ensure the components are separated by a single space."

"Example: IL-4 5000 U/ml + IL-2 100 U/ml + Anti-IFNγ 10 µg/ml + Anti-IL-12 10 µg/ml → 5000 U/ml + 100 U/ml + 10 µg/ml + 10 µg/ml."

"7. Per to Slash Conversion:\n"

" - Replace 'per' with '/'. For example, '0.2 ml per 25 g' should be '0.2 ml/25 g'.\n"

"8. Dosage Example Transformations:\n"

" - '0.5mg/ml Ascorbic Acid + 10mM β-Glycerol Phosphate' → '0.5 mg/ml + 10 mM'.\n"

" - '60% Calories' → '60% kcal'.\n"

" - '1X10^6' → '1 x 10^6'.\n"

" - '2 × 10^6 focal-forming units' → '2 × 10^6 focus forming units'.\n"

"- '50mg/kg/day + 5mg/kg/day + 20mg/kg/day + 60mg/kg/day' should be kept in its original format."

"- '10^7 PFUs' should be kept as '10^7 PFUs'."

"9. Standardizing Viral Particles:\n"

" - Use 'viral particles' consistently for clarity and precision.\n"

" - For combinations, preserve order and structure (e.g., '1 × 10^7 viral particles of MVA + 1 × 10^9 viral particles of adenovirus'→ '1 × 10^7 viral particles + 1 × 10^9 viral particles').\n\n"

"10. Consistency in Capitalization:\n"

" - Capitalize initial words but avoid capitalization for connecting words like 'and'.\n"

"11. Nutritional Context:\n"

" - Use consistent terminology for nutritional terms (e.g., '60% kcal' instead of '60% calories').\n"

" - If the context requires additional clarification, ensure terms align with standard nutritional guidelines.\n\n"

"12. Scientific Notation: Format: Use the format 'x x 10^n' for scientific notation."

"Examples:"

"'1.50E-04' → '1.5 x 10^-04'"

"'1.50E-05' → '1.5 x 10^-05'"

"'2.00E-04' → '2 x 10^-04'"

"For numbers like '1.10', '2.10', or '1.20', remove the trailing zero:"

"'1.5 E-04' → '1.5 x 10^-04'"

"'2.10E-05' → '2.1 x 10^-05'"

"13. Whole Numbers:Simplify Decimal Whole Numbers:"

"For numbers like '2.0' or '5.0', drop the '.0' and use the integer:"

"'2.0 mg/mL' → '2 mg/mL'"

"'5.0 µg' → '5 µg'"

"14. Units and Spacing: Leave a Space Between Numeric Value and Unit. Ensure a single space between the number and the unit."

"Examples:"

"'250μM + 200pM' → '250 µM + 200 pM'"

"'40mM' → '40 mM'"

"'2500nM + 100nM' → '2,500 nM + 100 nM'"

"'40ng/ml' → '40 ng/ml'"

"15. Commas in Large Numbers:Add a Comma for Readability:Insert commas for numbers ≥ '1,000'."

"Examples:"

"'10000 nM' → '10,000 nM'"

"'2000000 particles' → '2,000,000 particles'"

"'100000 U/mL' → '100,000 U/mL'"

"16. Keep a concise annotation only. Don't provide any explanations in any forms\n"

"Format the Output as a Table:\n"

"Create a table with two columns:\n"

"Column 1: List each original term separately.\n"

"Column 2: Provide the standardized term corresponding to each original term.\n\n"

"When formatting the output, please ensure that each label and its corresponding value are presented in plain text without any bolding or special characters like \*\*, -----. The output should look like this:"

"Output Format:\n"

" - Ensure all annotations follow a clean table format with two columns:\n"

" | Original Term | Standardized Term |\n"

" | 0.5mg/ml Ascorbic Acid | 0.5 mg/ml |\n"

" | 1X10^6 | 1 x 10^6 |\n"

" | MOI 3 | MOI 3 |\n"

" | 60% calories | 60% kcal |\n"

" | 2 × 10^6 focal-forming units | 2 × 10^6 focus forming units |\n"

" | 1 x 10^8 cfu | 1 x 10^8 CFU |\n"

" | µg/g body weight/day | µg/g bw/day |\n"

}

]