def create\_chatcompletion\_messages(row):

return [

{

"role": "system",

"content": "As a Biological Data Analyst, your primary task is to standardize the 'Model\_Type' field. This involves ensuring each model type is uniformly presented, clearly defined, and free from ambiguities or inconsistencies. You will review various biological model types, focusing on achieving standardization through precise guidelines."

},

{

"role": "user",

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

"Standardization Guidelines:\n"

"1. Capitalization Consistency: The initial word of every term should be capitalized. All other words should follow title case, except for specific technical notations (e.g., Greek letters).\n"

" Examples:\n"

" 'Allograft tail vein injection' → 'Allograft Tail Vein Injection'\n"

" 'Chemical Induced Disease Model: Acetaminophen induced Liver Injury' → 'Chemical Induced Disease Model: Acetaminophen Induced Liver Injury'\n"

"2. Spelling Out Greek Letters: Fully spell out Greek letters or symbols in terms.\n"

" Examples:\n"

" 'AdTGF-β1 induced Pulmonary Fibrosis' → 'AdTGFBeta Induced Pulmonary Fibrosis'\n"

" 'Conditional Knockout: Ezh2' → 'Conditional Knockout: Ezh2'\n"

"3. Short Form to Full Form: Expand all abbreviations or acronyms into their full forms unless the short form is more commonly recognized in scientific literature.\n"

" Examples:\n"

" 'CDX' → 'Cell Line-Derived Xenograft'\n"

" 'STZ' → 'Streptozotocin'\n"

"4. Uniform Terminology for Synonyms: Consolidate synonymous terms into a single standardized form.\n"

" Examples:\n"

" 'Knock-In' and 'Knockin' → 'Knockin'\n"

" 'Diabetes Mellitus' → 'Diabetes'\n"

"5. Format for Subcategories: Use the format 'Broad Category: Specifics' for clarity. Subcategories should be enclosed in parentheses when necessary.\n"

" Examples:\n"

" 'Chemical Induced Disease Model: STZ Induced Type 1 Diabetes Mellitus' → 'Chemical Induced Disease Model: Streptozotocin Induced Diabetes (Type 1)'\n"

" 'Chemical Induced Disease Model: Streptozotocin Induced Diabetic Cardiomyopathy' → 'Chemical Induced Disease Model: Streptozotocin Induced Diabetes (Cardiomyopathy)'\n"

"6. Consistency in Hyphenation and Spaces: Ensure consistent use of hyphens and spaces in compound words and phrases.\n"

" Examples:\n"

" 'N-Butyl-N-(4-hydroxybutyl) nitrosamine induced Bladder Cancer' → 'N-Butyl-N-(4-hydroxybutyl)nitrosamine Induced Bladder Cancer'\n"

" 'Chemical Induced Disease Model: AOM/DSS induced Colorectal Cancer' → 'Chemical Induced Disease Model: AOM-DSS Induced Colorectal Cancer'\n"

"7. Consolidation of Variants: Group variants of a term under a single standardized format where appropriate.\n"

" Examples:\n"

" 'Chemical Induced Disease Model: Bleomycin induced Pulmonary Fibrosis' → 'Chemical Induced Disease Model: Bleomycin Induced Pulmonary Fibrosis'\n"

" 'Chemical Induced Disease Model: DSS induced Colitis' → 'Chemical Induced Disease Model: Dextran Sulfate Sodium Induced Colitis'\n"

"8. Clarity in Broad Categories: Preserve the broad category (e.g., 'Chemical Induced Disease Model') in all cases to ensure clarity.\n"

" Examples:\n"

" 'Chemical Induced Disease Model: AOM DSS induced Colorectal Cancer' → 'Chemical Induced Disease Model: Azoxymethane-Dextran Sulfate Sodium Induced Colorectal Cancer'\n"

" 'Chemical Induced Disease Model: DSS induced Ulcerative Colitis' → 'Chemical Induced Disease Model: Dextran Sulfate Sodium Induced Ulcerative Colitis'\n"

"9. Retain Meaningful Specific Details: Keep details that contribute to scientific understanding, such as chemical names, while simplifying overly complex descriptions.\n"

" Examples:\n"

" 'Chemical Induced Disease Model: Alcohol Induced Liver Fibrosis' → 'Chemical Induced Disease Model: Ethanol Induced Liver Fibrosis'\n"

" 'Chemical Induced Disease Model: AOM Induced Colon Cancer' → 'Chemical Induced Disease Model: Azoxymethane Induced Colon Cancer'\n"

"10. Final Format and Review: Ensure each term adheres to the standardized format 'Broad Category: Specifics' and aligns with scientific conventions.\n"

" Examples:\n"

" 'Allograft orthotopic injection' → 'Allograft Orthotopic Injection'\n"

" 'Chemical Induced Disease Model: LPS induced Sepsis' → 'Chemical Induced Disease Model: Lipopolysaccharide Induced Sepsis'\n"

"\n"

"Final Format:\n"

"Create a table with two columns:\n"

"Column 1: List each original term exactly as it appears in the input. Preserve Original Term and do not apply any splitting, formatting, or transformations to the original input term. Simply copy it directly into the first column of the output.\n"

"Column 2: Provide the standardized term for each, ensuring clarity and consistency in how model types are reported across studies and publications.\n\n"

"Ensure accuracy and consistency based on scientific nomenclature standards and domain expertise. Do not add explanations in any format."

}

]