def create\_chatcompletion\_messages(row):

return [{"role": "system",

"content": "As a Biological Data Analyst, your primary task is to standardize and categorize the 'Revised RNA source' terms under their respective broader organ or tissue systems. Focus on the anatomical classification and ensure the exclusion of disease states from the categorization process."

},

{"role": "user",

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

"Task Overview:\n"

"You are provided with a list of biological terms, each representing a specific cell type, tissue, or cell line. Your task is to accurately categorize each term under the appropriate broader organ or tissue system based on anatomical classification only, omitting any references to diseases or pathological states.\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"

"Knowledge Application:\n"

"- Utilize specialized biological knowledge and authoritative databases to determine the origin and classification of complex cell lines, such as cancer cell lines known for tissue-specific origin.\n"

"- For cell lines like '4T1' and 'CT26', which are associated with breast and colon cancer respectively, ensure that the standardized categories reflect these specifics (e.g., '4T1' should be categorized under 'Mammary Gland' and 'CT26' under 'Colon').\n" "- Refer to sources such as the Cellosaurus Database for cell line specifics and recent literature for updates on cell line classifications.\n\n"

"Specific Instructions:\n"

"1. Comprehensive Source Review: Check multiple databases like Cellosaurus, Gene Ontology, and PubMed to cross-verify the tissue origin and typical use of the cell lines. Ensure all terms are standardized according to biological nomenclature standards and domain expertise."

"2. Detailed Contextual Categorization: Apply detailed context-specific knowledge, especially for terms that represent specific diseases or cancer types, to place them in the correct organ or tissue categories.\n"

"3. Clarify Ambiguities: Where classification is not straightforward or well-known, consult additional scientific publications or seek expert opinion when necessary.\n\n"

"4. Treat identical inputs with consistent standardized outputs, regardless of slight variations in the input term.\n\n"

"5. Focus on identifying the primary organ or tissue associated with each term. Use standardized terms that refer only to the anatomical location, such as 'Bone Marrow', 'Lymph Node', etc., and avoid adding any disease-related terms in your classifications.\n"

"6.For complex cell lines associated with specific diseases, such as 'BCR-ABL1 Murine ALL-like', classify them by their origin without mentioning the disease. For instance, classify 'BCR-ABL1 Murine ALL-like' as 'Immune System-Bone Marrow-NA'.\n"

"7. Ensure consistency by using the format 'System-Organ-Specific Part', with 'NA' used where specific parts are not applicable or the information is not available.\n\n"

"8. To accurately categorize the 'System' in biological terms, reference established classifications as below."

"Lymphatic System: Bone marrow, Spleen, Tonsils ,Lymph fluid, Lymph nodes, Lymph ducts, Lymph Vessels\n"

"Musculoskeletal System: Muscles, Bone,Cartilage,Ligaments,Tendons\n"

"Integumentary System: Skin,Hair,Nails,Subcutaneous tissue\n"

"Sensory System: Eye, Ear\n"

"Nervous System: Brain, Nerves, Spinal Cord\n"

"Female Reproductive System: Uterus, Ovary, Fallopian tube\n"

"Male Reproductive System: Testes, Prostate Gland\n"

"Urinary System: Kidney , Bladder,Urethra\n"

"Endocrine System: Thyroid Gland, Pituitary Gland, Adrenal Gland, Parathyroid Gland,Thymus,Pancreas\n"

"Digestive System: Mouth, Pharynx Esophagus,Stomach, Intestine, Liver, Gallbladder, Abdomen, Appendix, Colon,Salivary Gland,Palate\n"

"Cardiovascular System: Heart, Artery, Vein, Blood\n"

"Respiratory System: Lungs,Trachea, Nose\n"

"If it doesn’t fit in one of the categories, use ‘NA’.\n"

"Provide the table format with columns::\n"

"Create a table with two columns: Revised RNA Source Term, and Standardized Organ/Tissue. Do not provide any explanations in any form.\n"

"Column 1: List each given original Revised RNA Source term separately, exactly as given in the inputs.\n"

"Column 2: Provide the standardized organ/tissue category for each term, ensuring the format 'System-Organ-Specific Part' is followed, using 'NA' where information is not applicable or available.\n\n"

"Example Entries:\n"

"| Revised RNA Source Term | Standardized Organ/Tissue |\n"

"| Cell Line: 3T3 | Musculoskeletal System-Connective Tissue-NA |\n"

"| Cell Line: BCR-ABL1 Murine ALL-like | Immune System-Bone Marrow-NA |\n"

"| Cell Line: 4T1 | NA-Mammary Gland-NA |\n"

"| Cell Line: 4T1 Breast Cancer | NA-Mammary Gland-NA |\n"

"| Cell Line: CT26 | Digestive System-Colon-NA |\n"

"| Cell Line: CH12 | Immune System-NA-Lymphocyte |\n"

"| Cell Line: C2C12 | Musculoskeletal System-Muscle-Myoblast |\n"

"| Cnot3 cKO ESCs | NA-NA-Embryos |\n"

"| Cell Line: 16T | Endocrine System-Thyroid-NA |\n"

"| Cell Line: 3T3 | Musculoskeletal System-Connective Tissue-Adipose |\n"

"| Cell Line: 3T3 Tet-on 3G | Musculoskeletal System-Connective Tissue-Adipose |\n"

"| Cell Line: 3T3-F442A | Musculoskeletal System-Connective Tissue-Adipose |\n"

"| Cell Line: 3T3-L1 | Musculoskeletal System-Connective Tissue-Adipose |\n"

"| Cell Line: 3T3-L1 Adipocytes | Musculoskeletal System-Connective Tissue-Adipose |\n"

"| Cell Line: 3T3-L1 Pre-Adipocyte Connective Tissue| Musculoskeletal System-Connective Tissue-Adipose |\n"

}

]