Climate-Change-NER: Dataset information

December 14, 2024

This supplementary document provides the definitions of the 13 named entity categories contained in the dataset Climate-Change-NER. The definitions have been obtained from the dataset card available on Hugging Face.[1] In addition to the category definitions that have been used to build the prompts, Table 1 gives an overview of the frequency of NE instances per NE class and per data split. Tables 2 and 3 give basic statistical information about the average, median, maximum, and minimum length of sentences and tokens in each data split and in the whole dataset, alongside the total number of sentences and tokens per data split and in the complete dataset. Finally, Tables 4 and 5 list the ten most frequent and least frequent real-world instances of the 13 NE categories respectively. This data is presented per data split. For weakly represented NE classes, such as the class *CLIMATE-GREENHOUSE-GASES* in the development data split, the number of NE instances in the two tables is lower than ten, due to the fact that the total number of instances is less than ten.[2] The NE frequencies presented here are raw counts.

Definitions of NE classes

- **CLIMATE-HAZARDS**: hazards with potential negative impact on climate, such as floods, wildfires, droughts, and heatwaves. Where a hazard is named in more detail in a text, the entire term is annotated, e.g., surface water flood or soil liquefaction;

- **CLIMATE-MITIGATIONS**: activities to reduce climate change or to better deal with the consequences;

- **CLIMATE-PROPERTIES**: properties of the climate itself (not abstract objects like models and datasets) that typically come with values and units;

- **CLIMATE-NATURE**: aspects of nature that are not alive, such as oceans, rivers, the atmosphere, winds, and snow;

- **CLIMATE-MODELS**: specific physical, mathematical, or artificial intelligence objects, nowadays always computer-executable, used to analyze and usually predict climate parameters;

- **CLIMATE-PROBLEM-ORIGINS**: problems that describe why the climate is changing. Key examples are fossil fuel and deforestation. We also mention sectors that can be cited as causes of energy use. For instance, in a text about the energy consumption by the transport sector, transport sector is annotated as problem;

- **CLIMATE-OBSERVATIONS**: climate observation tools with a name. Examples are satellites, radiospectrometers, rain gauges, wildlife cameras, and questionnaires;

---

[1] https://huggingface.co/datasets/ibm/Climate-Change-NER

[2] The NE instance *CO2* is the only NE instance of this category in the development dataset, with a total count of 2.

- **CLIMATE-ASSETS**: objects or services of value to humans that can get destroyed or diminished by climate-hazards. Key categories are health, buildings, infrastructure, and crops or livestock;

- **CLIMATE-IMPACTS**: effects of hazards, primarily negative effects on humans. We also consider impacts on livestock as impacts, as it indirectly affects humans;

- **CLIMATE-GREENHOUSE-GASES**: gases that cause heating of the atmosphere (greenhouse gases);

- **CLIMATE-ORGANIZATIONS**: real-world organizations with climate-related interests;

- **CLIMATE-ORGANISMS**: animals, plants, and other organisms that are considered for their own sakes (in contrast to as food for humans) as climate organisms;

- **CLIMATE-DATASETS**: specific collections of climate data with a name. A climate dataset can be the result of observations or of a model, e.g., as a prediction or reanalysis. The data may be lists, tables, databases, inventories or historical records, where the data dominate over attached code.

Named entity class	Train	Dev	Test	Total per category
CLIMATE-HAZARDS	320	50	34	404
CLIMATE-MITIGATIONS	185	30	38	253
CLIMATE-PROPERTIES	455	107	86	648
CLIMATE-NATURE	705	195	98	998
CLIMATE-MODELS	325	78	94	497
CLIMATE-PROBLEM-ORIGINS	129	19	20	168
CLIMATE-OBSERVATIONS	105	4	21	130
CLIMATE-ASSETS	248	31	50	329
CLIMATE-IMPACTS	63	16	17	96
CLIMATE-GREENHOUSE-GASES	25	2	31	58
CLIMATE-ORGANIZATIONS	112	35	30	177
CLIMATE-ORGANISMS	203	17	11	231
CLIMATE-DATASETS	154	28	25	207
Total per data split	3029	612	555	4196

Table 1: Named entity instances per category and per data split in Climate-Change-NER

Data split	Count	Average len.	Median len.	Maximum len.	Minimum len.
train	985	32	29	115	2
development	191	33.04	30	86	1
test	177	32.63	31	97	10
total	1353	32.23	30	115	2

Table 2: Climate-Change-NER sentence features

Data split	Count	Average len.	Median len.	Maximum len.	Minimum len.
train	31516	4.91	4	21	1
development	6311	4.89	4	18	1
test	5775	4.91	4	21	1
total	43602	4.9	4	21	1

Table 3: Climate-Change-NER token features

Most frequent NE instances per category in Climate-Change-NER			
NE Class	**TRAIN**	**DEVELOPMENT**	**TEST**
CLIMATE-HAZARDS	flood, fire, drought, fires, sea level rise, floods, landslides, pollution, storm surge, extinction	drought, pollution, forest fire, flood, water scarcity, biomass burning, forest fires, climate extremes, Ixodes scapularis, tick-borne pathogens	flood, fire, SLR, Fires, drought, soil moisture depletion, fires, hurricanes, earthquakes, storm surges
CLIMATE-MITIGATIONS	irrigation, renewable energy, mitigation, climate policy, electric vehicles, carbon tax, FMNR, nuclear power, SAT, Dam	eco-environmental management, insulation, Wind energy, mitigation, water use efficiency, water savings, climate-smart agricultural, CSA, greenhouse gas mitigation policies, Biosphere Reserve	irrigation, urban water management, greenhouse gas abatement, irrigated, urban irrigation, anaerobic digestion, climate regulations, natural gas-fired combined cycle, NGCC, Clean Air Interstate Rule
CLIMATE-PROPERTIES	temperature, streamflow, precipitation, burned area, discharge, soil moisture, altitude, Precipitation, solar radiation, albedo	temperature, precipitation, SOS, OHC, wind speed, soil moisture, SPEI, GPP, ZTD, EOS	temperature, soil moisture, ρ_{eff}, ETref, effective density, EE, thermal comfort
CLIMATE-NATURE	precipitation, vegetation, rainfall, aerosol, hydrological, forest, runoff, ocean, atmosphere, aerosols	precipitation, atmosphere, reef, aerosol, glacier, atmospheric, ocean, land surface, sea ice, SAOD	precipitation, rainfall, atmosphere, tundra, aerosols, water balance, snow, water vapor, urban vegetation, ET
			Continued on next page

NE CLASS	TRAIN	DEVELOPMENT	TEST
CLIMATE-MODELS	SWAT, DSSAT, HBV, CCSM, MAR, WRF, NARCCAP, CMIP3, HadCM3, CERES-Rice	CMIP5, CMIP6, CMIP3, WRF, LDAS-Monde, PCR-GLOBWB, RACMO, HIRHAM, Integrated Valuation of Ecosystem Services and Tradeoffs, System of Integrated Environmental and Economic Accounting	CMIP5, STIRPAT, RegCM4, EMIL, WRF-UCM, DCSM, Whole Atmosphere Community Climate Model, WACCM, EPIC, Coupled Model Intercomparison Project
CLIMATE-PROBLEM-ORIGINS	emission, emissions, fossil fuel, urbanization, LUCC, land use change, fossil fuels, population growth, land use changes, NG	emission, land use change, urbanization, LUC, nutrient loading, toxic substances, water abstraction, population growth, corruption, food production	emission, manure, fossil fuel combustion, emissions, population growth, Manure, intestines of animals, coal, fuel wood, space heating
CLIMATE-OBSERVATIONS	NDVI, MODIS, Landsat, lidar, ALOS, PAL-SAR, HAMSR, ALS, GPS, ERBE	Landsat, research cruises SO234-2, SO235, RV SONNE	SAR, TerraSAR-X, Sentinel-1, campaign STABLE, Beijing Station, DMA-CPMA-SP2, SP2, Lyman–Birge–Hopfield, LBH, band systems
CLIMATE-ASSETS	agriculture, agricultural, crop, livestock, food security, wheat, building, maize, health, food	water resources, agricultural, food security, farmers, eco-environmental benefits, water supplies, farm, crop yields, water infrastructure, health benefits	agricultural, livestock, nutrients, water availability, crop, buildings, health, monocrops, maize, urban water systems
CLIMATE-IMPACTS	damage, damages, diseases, disease, disaster, deaths, pneumonia, mortality, disruptions, downy mildew	Lyme disease, poverty, malaria, eco-livelihood impacts, economic losses, encephalitis, babesiosis, anaplasmosis, homeless, disruption	climate catastrophe, continuous damage, killing, Burn area, plant diseases, Food insecurity, poverty, unsustainable livelihoods, loss of crop, exhaustion
			Continued on next page

NE CLASS	TRAIN	DEVELOPMENT	TEST
CLIMATE-GREENHOUSE-GASES	CO2, carbon dioxide, methane, BC, CH4, Carbon dioxide, Non-Methane Hydrocarbons, NMHC, NOx	CO2	methane, carbon dioxide, CO2, rBC, perfluorocarbons, PFCs, decafluorobutane, C4F10, dodecafluoropentane, C5F12
CLIMATE-ORGANIZATIONS	IPCC, ECMWF, NCEP, GMEP, Geophysical Fluid Dynamics Laboratory, BRI, IIASA, GFDL, NASA, WOCE	SALKKU, IPCC, SMHI, Expert Team on Climate Change Detection and Indices, ETCCDI, National Oceanic and Atmospheric Administration Geophysical Fluid Dynamics Laboratory, Interreg IVB project AMICE, FEM, UNESCO, European Commission	IPCC, North American Regional Climate Change Assessment Program, NASA, DAMOCLES, Australian Bureau of Meteorology, CIERA, European Centre for Medium-Range Weather Forecasts, ECMWF, Climatic Research Unit, CRU
CLIMATE-ORGANISMS	species, habitat, biodiversity, plant, frogs, phytoplankton, tree, Joshua trees, diversity, butterfly	species, pine, Ecosystem diversity, rare species, snow leopard, demersal fish, biodiversity, birds, vascular plants, endemic	plant, eelgrass, biological species, habitat, Bryophytes, habitats, trees, Zostera marina L.
CLIMATE-DATASETS	TRMM, A2, CMORPH, UTCI, RCP8.5, SRES, TMPA, A1B, Tropical Rainfall Measuring Mission, ERA-Interim	CAMS-OPI, AR5, APHRODITE, Princeton Global Forcing, Soil-adjusted Vegetation Index, SAVI, Index-based Built-up Index, IBI, Soil Brightness Index, NDSI	RCP4.5, RCP8.5, A2, CIMIS, B2, ERA-Interim 6, ERA-Interim, CRU, Fourth Assessment Report, Climate Hazards group InfraRed Precipitation with Station data

Table 4: Ten most frequent NE instances in each category for every data split of Climate-Change-NER

Least frequent NE instances per category in Climate-Change-NER			
NE Class	**TRAIN**	**DEVELOPMENT**	**TEST**
CLIMATE-HAZARDS	Tropical Cyclones, Extratropical Transition, extratropical transition, ET, cyclone, Tropical Cyclone, Lightning, LANDSLIDE, volcano, debris flowing	shoreline instability, sea level rise, Water scarcity, water gap, fire, PM2.5 emissions, fire emissions, ozone depletion, Flood, extreme precipitation	woody plant encroachment, insect pests, ocean acidification, pests, Sea level rise, nitrate leaching, fire emissions, Drought, Inundation, flooding
CLIMATE-MITIGATIONS	National Plans to Combat Desertification, NPCD, cycling, alternative cropping systems, Urban water management, stormwater capture, strategy for adaptation, alternative management strategies, fire risk prevention, Agroforestry	soil conservation technique, residue mulch, pigeon pea hedges, wind technology, wind energy, trading scheme, fire management, mitigation strategy, water recycling, disaster preparedness	early warning systems, disaster relief strategies, building envelopes, double facades, solar chimney, passive and active solar control systems, wildlife reservoirs, clothing insulation, Paris Agreement, suppression of fire
CLIMATE-PROPERTIES	dNBR, aspect, damage, SPEI, WCI, PRECIPITATION, population, surface displacement, oxygen lines, water - vapor line	form drag, sea level changes, daily precipitation, air temperature, atmospheric lifetimes, oceanic and atmospheric concentrations, tropical lifetime, Standardized Precipitation Index, SPI, discharge	age, exposure to sun, population dynamics, sedimentation, ice - impact rates, equilibrium temperature, particle size, sea - ice cover, Precipitation, precipitations
CLIMATE-NATURE	tropical island, convective, stratiform, rainclouds, raincloud, tectonic plates, tropical, stream, Rainstorms, rainstorm	ice cover, sea level, Coastal, stratosphere, oceanic emissions, Asian monsoon anticyclone, oceanic, Water Resources, watershed, coast	winds, aerosol, clouds, sea - ice retreat, dust plume, Rainfall, water - resource, hydroclimatic, soil, catchment

NE CLASS	TRAIN	DEVELOPMENT	TEST
CLIMATE-MODELS	HadGEM2, RAC-MOv2, CMIP6, seasonal climate forecast system, SEAS5, Fire Events Delineation, FSU superensemble, UKMO, CPTEC, Global Metropolitan Detector	FireMIP, Community Land Model, CLM, Joint UK Land environment Simulator - Interactive Fire And Emission Algorithm For Natural Environments, JULES - INFERNO, Advanced Weather Research and Forecasting, WRF - ARW, MED - CORDEX, MikeShe, Mike11	CARMA, pSIMS, AP-SIM, DSSAT, INteractive Fires and Emissions algoRithm for Natural environments, UK 's Earth System Model, UKESM1, Waterdyn, Australian Water Resource Assessment, AWRA - L
CLIMATE-PROBLEM-ORIGINS	impervious areas, demographic pressure, landuse / land cover changes, land - use, Natural Gas, hydraulic fracturing, coal, natural gas, power plants, power plant	nutrient loading, toxic substances, water abstraction, population growth, corruption, food production, Inadequate timber extraction, cattle, abusive recreational practices, urban expansion	Manure, intestines of animals, coal, fuel wood, space heating, heating, gas and electricity consumption, human damage activities, urbanization, anthropogenic ignition
CLIMATE-OBSERVATIONS	Differential Interferometric of Synthetic Aperture Radar, Global Positioning System, High - Altitude MMIC Sounding Radiometer, High - Altitude Monolithic Microwave Integrated Circuit (MMIC) Sounding Radiometer, CAMEX-4, Tropical Cloud Systems and Processes, African Monsoon Multidisciplinary Analyses, GH, Synthetic Aperture Radar, TOPEX / POSEIDON Radar Altimeter	Landsat, research cruises SO234 - 2, SO235, RV SONNE	Lyman – Birge – Hopfield, LBH, band systems, Analytical Spectral Device (ASD) Field Spec Pro, Landsat 8 Operational Land Imager, OLI, SALTRACE, lidar, C- and X - band, C - band

NE CLASS	TRAIN	DEVELOPMENT	TEST
CLIMATE-ASSETS	croplands, human health, FOOD SECURITY, urban areas, healthy diets, fruit, income opportunities, Legumes, forages, drinking water	national welfare, forestry, Transport infrastructure networks, infrastructure, vehicle, cassava, Smallholder, food supply, crops, Water supply	income, smallholder, high - rise, skyscrapers, built environments, wellbeing, pea, oat, soybean, agricultural productivity
CLIMATE-IMPACTS	flood damage, disruption, detrimental, personal losses, Japanese encephalitis, bovine tuberculosis, famine, homeless, food insecurity, disastrous	eco - livelihood impacts, economic losses, encephalitis, babesiosis, anaplasmosis, homeless, disruption, disruptions, flood footprint, traffic disruptions	unsustainable livelihoods, loss of crop, exhaustion, illness, destruction, calamities, Bovine tuberculosis, zoonosis, mortality, disaster
CLIMATE-GREENHOUSE-GASES	CO2, carbon dioxide, methane, BC, CH 4, Carbon dioxide, CO 2, Non - Methane Hydrocarbons, NMHC, NOx	CO 2	C4F10, dodecafluoropentane, C 5F12, tetradecafluorohexane, C6F14, hexadecafluoroheptane, C 7F16, octadecafluorooctane, C 8F18, black carbon
CLIMATE-ORGANIZATIONS	CMA, CMC, BOM, MF, KMA, eartH2Observe, NOAA, Vaisala, Jet Propulsion Laboratory, JPL	AfriCultuReS, CNRM, Centre National de Recherches Météorologiques, European Center for Medium Range Weather Forecast, ECMWF, SEAREG, Swedish Meteorological and Hydrological Institute, APEC Climate Center, APCC, ENSEMBLES	BMD, Australian Water Availability Project, AWAP, Scaling and Assimilation of Soil Moisture and Streamflow, SASMAS, ZKI, Center for Satellite - Based Crisis Information, German Aerospace Center, DLR, European Space Agency
CLIMATE-ORGANISMS	biosphere, indigenous animals, white - tailed deer, Rare plant, Trifolium repens L., Trifolium vesiculosum Savi, Clover, Loblolly Pine, Pinus taeda L., Tree	rare species, snow leopard, demersal fish, biodiversity, birds, vascular plants, endemic, organisms, plant species, habitat	plant, eelgrass, biological species, habitat, Bryophytes, habitats, trees, Zostera marina L.
			Continued on next page

NE CLASS	TRAIN	DEVELOPMENT	TEST
CLIMATE-DATASETS	Precipitation Estimation from Remotely Sensed information using Artificial Neural Networks, FAOSTAT, EarthStat, WorldClim, National Lightning Detection Network, Long Range Lightning Detection Network, Canadian Lightning Detection Network, CLDN, Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks, Integrated Multi - satellitE Retrievals for Global	Global Fire Emissions Database, GFED, ERA - INTERIM, Climate Assessment and Dataset, ECA&D, ERA - Interim, UK Foresight Future Flooding Report, Climate Anomaly Monitoring System - Outgoing Longwave Radiation Precipitation Index, Asian Precipitation - Highly - Resolved Observational Data Integration Towards Evaluation, A1B	CHIRPS, E - OBS, Climate Prediction Center MORPHING, CMORPH, Tropical Rainfall Measuring Mission, TRMM, Precipitation Estimation Algorithm from Remotely - Sensed Information using an Artificial Neural Network, PERSIANN, global Satellite Mapping of Precipitation, GSMaP

Table 5: Ten least frequent NE instances in each category for every data split of Climate-Change-NER