

model_03a_bert_base_triplet_margin

May 4, 2023

1 Model 03a

Evidence retrieval using a Siamese BERT classification model. This is similar to Model 01, however, it only uses official pre-trained models from hugging face.

Ref: - [Hugging face pre-trained models](#) - [Hugging face guide to fine-tuning](#) - [Hugging face guide to fine-tuning easy](#) - [SO Guide](#)

1.1 Setup

1.1.1 Working Directory

```
[ ]: # Change the working directory to project root
from pathlib import Path
import os
ROOT_DIR = Path.cwd()
while not ROOT_DIR.joinpath("src").exists():
    ROOT_DIR = ROOT_DIR.parent
os.chdir(ROOT_DIR)
```

1.1.2 File paths

```
[ ]: MODEL_PATH = ROOT_DIR.joinpath("./result/models/*")
DATA_PATH = ROOT_DIR.joinpath("./data/*")
NER_PATH = ROOT_DIR.joinpath("./result/ner/*")
```

1.1.3 Dependencies

```
[ ]: # Imports and dependencies
import torch
from torch.utils.data import Dataset, DataLoader
from torch.nn import Module, TripletMarginWithDistanceLoss
from transformers import BertModel, BertTokenizer
from torch.optim import Adam
from torch.optim.lr_scheduler import LinearLR
from torcheval.metrics import BinaryAccuracy, BinaryF1Score

from src.torch_utils import get_torch_device
```

```

import json
from dataclasses import dataclass
from typing import List, Union, Tuple
from tqdm import tqdm
import random
import numpy as np
from datetime import datetime
from math import exp

TORCH_DEVICE = get_torch_device()

```

Torch device is 'mps'

1.2 Dataset

```

[ ]: @dataclass
class ClaimEvidencePair:
    claim_id:str
    evidence_id:str
    label:int = 0

```

```

[ ]: @dataclass
class ClaimEvidenceTriple:
    claim_id:str
    pos_evidence_id:str
    neg_evidence_id:str

```

```

[ ]: class SiameseEvalDataset(Dataset):
    def __init__(
        self,
        dev_claims_path:Path,
        evidence_path:Path,
        device = None,
        verbose:bool=True
    ) -> None:
        super(SiameseEvalDataset, self).__init__()
        self.verbose = verbose
        self.device = device

        # Load claims data from json
        with open(dev_claims_path, mode="r") as f:
            self.claims = (json.load(fp=f))

        # Load evidence library
        self.evidence = dict()
        with open(evidence_path, mode="r") as f:
            self.evidence.update(json.load(fp=f))

```

```

    # Get a list of all evidences within the dev set
    self.related_evidences = sorted({
        evidence_id
        for claim in self.claims.values()
        for evidence_id in claim["evidences"]
    })

    # Generate the data
    self.data = self.__generate_data()
    return

def __generate_data(self):
    data = []
    for claim_id, claim in tqdm(
        iterable=self.claims.items(),
        desc="claims",
        disable=not self.verbose
    ):
        evidence_ids = claim["evidences"]

        # Get the positives
        for evidence_id in evidence_ids:
            data.append(ClaimEvidencePair(
                claim_id=claim_id,
                evidence_id=evidence_id,
                label=1
            ))

        # Get some negatives
        n_neg = 0
        for rel_evidence_id in self.related_evidences:
            if n_neg >= 10:
                break
            if rel_evidence_id in evidence_ids:
                continue
            data.append(ClaimEvidencePair(
                claim_id=claim_id,
                evidence_id=rel_evidence_id,
                label=-1
            ))
            n_neg += 1
        return data

def __len__(self):
    return len(self.data)

```

```

def __getitem__(self, idx) -> Tuple[Union[str, torch.Tensor]]:
    # Fetch the required data rows
    data = self.data[idx]

    # Get the label
    label = torch.tensor(data.label, device=self.device)

    # Get text ids
    claim_id = data.claim_id
    evidence_id = data.evidence_id

    # Get text
    claim_text = self.claims[claim_id]["claim_text"]
    evidence_text = self.evidence[evidence_id]

    return (claim_text, evidence_text, label)

```

```

[ ]: class SiameseTripletDataset(Dataset):

    def __init__(
        self,
        claims_paths:List[Path],
        claims_shortlist_paths:List[Path],
        evidence_path:Path,
        evidence_shortlists:List[Path] = None,
        device = None,
        n_neg_shortlist:int = 10,
        n_neg_general:int = 10,
        verbose:bool=True
    ) -> None:
        super(SiameseTripletDataset, self).__init__()
        self.verbose = verbose
        self.device = device
        self.n_neg_shortlist = n_neg_shortlist
        self.n_neg_general = n_neg_general

        # Load claims data from json, this is a list as we could use
        # multiple json files in the same dataset
        self.claims = dict()
        for json_file in claims_paths:
            with open(json_file, mode="r") as f:
                self.claims.update(json.load(fp=f))
                # print(f"loaded claims: {json_file}")

        # Load the pre-retrieved shortlist of evidences by claim
        self.claims_shortlist = dict()
        for json_file in claims_shortlist_paths:

```

```

        with open(json_file, mode="r") as f:
            self.claims_shortlist.update(json.load(fp=f))
            # print(f"loaded claims_shortlist: {json_file}")

    # Load evidence library
    self.evidence = dict()
    with open(evidence_path, mode="r") as f:
        self.evidence.update(json.load(fp=f))
        # print(f"loaded evidences: {json_file}")

    # Load the evidence shortlists if available
    # Reduce the overall evidence list to the shortlist
    if evidence_shortlists is not None:
        self.evidence_shortlist = set()
        for json_file in evidence_shortlists:
            with open(json_file, mode="r") as f:
                self.evidence_shortlist.update(json.load(fp=f))
                # print(f"loaded evidence shortlist: {json_file}")

    # print(f"n_evidences: {len(self.evidence)}")

    # Generate the data
    self.data = self.__generate_data()
    return

def __generate_data(self):
    print("Generate siamese dataset")

    data = []
    for claim_id, claim in tqdm(
        iterable=self.claims.items(),
        desc="claims",
        disable=not self.verbose
    ):
        pos_evidence_ids = set(claim["evidences"])
        shortlist_neg_evidence_ids = \
            set(self.claims_shortlist.get(claim_id, [])) \
            .difference(pos_evidence_ids)
        general_neg_evidence_ids = self.evidence_shortlist

        # For each positive evidence
        for pos_evidence_id in pos_evidence_ids:

            # Add i number of negatives from shortlisted evidences
            for i in range(self.n_neg_shortlist):
                if len(shortlist_neg_evidence_ids) > 0:
                    neg_evidence_ids = random.sample(

```

```

        population=shortlist_neg_evidence_ids,
        k=min(
            self.n_neg_shortlist,
            len(shortlist_neg_evidence_ids)
        )
    )

    for neg_evidence_id in neg_evidence_ids:
        data.append(ClaimEvidenceTriple(
            claim_id=claim_id,
            pos_evidence_id=pos_evidence_id,
            neg_evidence_id=neg_evidence_id
        ))

    # Add j number of negatives from general evidences
    for j in range(self.n_neg_general):
        neg_evidence_ids = random.sample(
            population=general_neg_evidence_ids,
            k=min(self.n_neg_general, len(general_neg_evidence_ids))
        )

        for neg_evidence_id in neg_evidence_ids:
            data.append(ClaimEvidenceTriple(
                claim_id=claim_id,
                pos_evidence_id=pos_evidence_id,
                neg_evidence_id=neg_evidence_id
            ))

        continue

    print(f"Generated data n={len(data)}")

    return data

def __len__(self):
    return len(self.data)

def __getitem__(self, idx) -> Tuple[str]:
    # Fetch the required data rows
    data = self.data[idx]

    # Get text
    anchor_text = self.claims[data.claim_id]["claim_text"]
    pos_evidence_text = self.evidence[data.pos_evidence_id]
    neg_evidence_text = self.evidence[data.neg_evidence_id]

    return (anchor_text, pos_evidence_text, neg_evidence_text)

```

```
[ ]: # WE WILL GENERATE THE DATASET PER EPOCH SO TO RANDOMISE THE NEGATIVE SAMPLES

# train_data = SiameseTripletDataset(
#     claims_paths=[DATA_PATH.with_name("train-claims.json")],
#     claims_shortlist_paths=[NER_PATH.
# ↪with_name("train_claim_evidence_retrieved.json")],
#     evidence_shortlists=[NER_PATH.
# ↪with_name("shortlist_train_claim_evidence_retrieved.json")],
#     evidence_path=DATA_PATH.with_name("evidence.json"),
#     device=TORCH_DEVICE,
#     n_neg_shortlist=2,
#     n_neg_general=1
# )

[ ]: # for i in range(50):
#     anchor, pos, neg = train_data[i]
#     print(anchor[:50], pos[:50], neg[:50])
```

1.3 Build model

```
[ ]: class SiameseTripletEmbedderBert(Module):

    def __init__(
        self,
        pretrained_name:str,
        device,
        **kwargs
    ) -> None:
        super(SiameseTripletEmbedderBert, self).__init__(**kwargs)
        self.device = device

        # Use a pretrained tokenizer
        self.tokenizer = BertTokenizer.from_pretrained(pretrained_name)

        # Use a pretrained model
        self.bert = BertModel.from_pretrained(pretrained_name)
        self.bert.to(device=device)
        return

    def forward(self, anchor_texts, pos_texts, neg_texts=None) -> Tuple[torch.
    ↪Tensor]:

        # Run the tokenizer
        t_kwargs = {
            "return_tensors": "pt",
            "padding": True,
            "truncation": True,
```

```

        "max_length": 100,
        "add_special_tokens": True
    }
    anchor_x = self.tokenizer(anchor_texts, **t_kwargs)
    pos_x = self.tokenizer(pos_texts, **t_kwargs)
    if neg_texts:
        neg_x = self.tokenizer(neg_texts, **t_kwargs)

    anchor_x = anchor_x["input_ids"].to(device=self.device)
    pos_x = pos_x["input_ids"].to(device=self.device)
    if neg_texts:
        neg_x = neg_x["input_ids"].to(device=self.device)

    # Run Bert
    anchor_x = self.bert(anchor_x, return_dict=True).pooler_output
    pos_x = self.bert(pos_x, return_dict=True).pooler_output
    if neg_texts:
        neg_x = self.bert(neg_x, return_dict=True).pooler_output
    # dim=768

    if neg_texts:
        return anchor_x, pos_x, neg_x
    else:
        return anchor_x, pos_x

```

1.4 Training and evaluation loop

```

[ ]: model = SiameseTripletEmbedderBert(
    pretrained_name="bert-base-cased",
    device=TORCH_DEVICE
)

```

Some weights of the model checkpoint at bert-base-cased were not used when initializing BertModel: ['cls.predictions.decoder.weight', 'cls.predictions.transform.LayerNorm.bias', 'cls.seq_relationship.weight', 'cls.seq_relationship.bias', 'cls.predictions.bias', 'cls.predictions.transform.dense.weight', 'cls.predictions.transform.LayerNorm.weight', 'cls.predictions.transform.dense.bias']

- This IS expected if you are initializing BertModel from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a BertForSequenceClassification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing BertModel from the checkpoint of a model that you expect to be exactly identical (initializing a BertForSequenceClassification model from a BertForSequenceClassification model).


```
[ ]: loss_fn = TripletMarginWithDistanceLoss(
    # distance_function=torch.cdist, # Euclidean distance, not implemented for MPS
    # Default is Pairwise Distance
    margin=5
)
optimizer = Adam(
    params=model.parameters(),
    lr=0.0000002
) #! Hyperparams
```

```
[ ]: run_time = datetime.now().strftime('%Y_%m_%d_%H_%M')
MODEL_NAME = f"model_03a_bert_base_triplet_{run_time}.pth"
N_EPOCHS = 100
BATCH_SIZE = 64
```

```
[ ]: dev_data = SiameseEvalDataset(
    dev_claims_path=DATA_PATH.with_name("dev-claims.json"),
    evidence_path=DATA_PATH.with_name("evidence.json"),
    device=TORCH_DEVICE
)

dev_dataloader = DataLoader(
    dataset=dev_data,
    shuffle=False,
    batch_size=BATCH_SIZE
)
```

```
claims: 100%|          | 154/154 [00:00<00:00, 178234.77it/s]
claims: 100%|          | 154/154 [00:00<00:00, 178234.77it/s]
```

```
[ ]: import warnings
warnings.filterwarnings('ignore')
```

```
[ ]: # Run evaluation before training to establish baseline
model.eval()

dev_batches = tqdm(dev_dataloader, desc="dev batches")
epoch_pos_cos_sim = []
epoch_neg_cos_sim = []
for batch in dev_batches:
    claim_texts, evidence_texts, labels = batch

    # Forward
    claim_emb, evidence_emb = model(claim_texts, evidence_texts)

    # Cosine similarity
```

```

cos_sim = torch.cosine_similarity(claim_emb, evidence_emb)
labelled_cos_sim = cos_sim * labels
pos_cos_sim = labelled_cos_sim[torch.where(labelled_cos_sim > 0)]
neg_cos_sim = labelled_cos_sim[torch.where(labelled_cos_sim < 0)]

batch_pos_cos_sim = torch.mean(pos_cos_sim).cpu().item()
batch_neg_cos_sim = torch.mean(neg_cos_sim).cpu().item() * -1

epoch_pos_cos_sim.append(batch_pos_cos_sim)
epoch_neg_cos_sim.append(batch_neg_cos_sim)

dev_batches.postfix = f"pos cos_sim: {batch_pos_cos_sim:.3f}" + \
    f" neg cos_sim: {batch_neg_cos_sim:.3f}"

continue

```

```

dev batches: 100%|          | 32/32 [00:07<00:00, 4.22it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]

```

```

[ ]: print(f"Average cos sim (pos, neg): {np.mean(epoch_pos_cos_sim):3f}, {np.
      ↪mean(epoch_neg_cos_sim):3f}")

```

```

Average cos sim (pos, neg): 0.863419, 0.838396

```

```

[ ]: metric_accuracy = BinaryAccuracy()
metric_f1 = BinaryF1Score()
metric_recall = BinaryF1Score()

scheduler = LinearLR(
    optimizer=optimizer,
    start_factor=0.1,
    end_factor=1,
    total_iters=BATCH_SIZE/10,
    verbose=True
)
best_epoch_loss = 999
for epoch in range(N_EPOCHS):

    print(f"Epoch: {epoch} of {N_EPOCHS}\n")

    # Run training
    model.train()

    train_data = SiameseTripletDataset(
        claims_paths=[DATA_PATH.with_name("train-claims.json")],
        claims_shortlist_paths=[NER_PATH.
    ↪with_name("train_claim_evidence_retrieved.json")],

```

```

        evidence_shortlists=[NER_PATH.
↪with_name("shortlist_train_claim_evidence_retrieved.json")],
        evidence_path=DATA_PATH.with_name("evidence.json"),
        device=TORCH_DEVICE,
        n_neg_shortlist=2,
        n_neg_general=1
    )

    train_dataloader = DataLoader(
        dataset=train_data,
        shuffle=True,
        batch_size=BATCH_SIZE
    )

    train_batches = tqdm(train_dataloader, desc="train batches")
    running_losses = []
    for batch in train_batches:
        anchor, positive, negative = batch

        # Reset optimizer
        optimizer.zero_grad()

        # Forward + loss
        anchor_emb, pos_emb, neg_emb = model(anchor, positive, negative)
        loss = loss_fn(anchor=anchor_emb, positive=pos_emb, negative=neg_emb)

        # Backward + optimiser
        loss.backward()
        optimizer.step()

        # Update running loss
        batch_loss = loss.item() * len(batch)
        running_losses.append(batch_loss)

        train_batches.postfix = f"loss: {batch_loss:.3f}"

        continue

    scheduler.step()

    epoch_loss = np.average(running_losses)
    print(f"Average epoch loss: {epoch_loss}")

    # Save model
    if epoch_loss <= best_epoch_loss:
        best_epoch_loss = epoch_loss
        torch.save(model, MODEL_PATH.with_name(MODEL_NAME))

```

```

    print(f"Saved model to: {MODEL_PATH.with_name(MODEL_NAME)}")

    # Evaluate every 5 epochs
    # if epoch % 5 != 0:
    #     continue

    # Run evaluation before training to establish baseline
    model.eval()

    dev_batches = tqdm(dev_dataloader, desc="dev batches")
    epoch_pos_cos_sim = []
    epoch_neg_cos_sim = []
    for batch in dev_batches:
        claim_texts, evidence_texts, labels = batch

        # Forward
        claim_emb, evidence_emb = model(claim_texts, evidence_texts)

        # Cosine similarity
        cos_sim = torch.cosine_similarity(claim_emb, evidence_emb)
        pos_cos_sim = labelled_cos_sim[torch.where(labelled_cos_sim > 0)]
        neg_cos_sim = labelled_cos_sim[torch.where(labelled_cos_sim < 0)]

        batch_pos_cos_sim = torch.mean(pos_cos_sim).cpu().item()
        batch_neg_cos_sim = torch.mean(neg_cos_sim).cpu().item() * -1

        epoch_pos_cos_sim.append(batch_pos_cos_sim)
        epoch_neg_cos_sim.append(batch_neg_cos_sim)

        dev_batches.postfix = f"pos cos_sim: {batch_pos_cos_sim:.3f}" + \
            f" neg cos_sim: {batch_neg_cos_sim:.3f}"

    continue

    print(f"Average cos sim (pos, neg): {np.mean(epoch_pos_cos_sim):3f}, {np.
    ↪mean(epoch_neg_cos_sim):3f}")

print("Done!")

```

Adjusting learning rate of group 0 to 2.0000e-08.
Epoch: 0 of 100

Generate siamese dataset

claims: 100%| | 1228/1228 [00:23<00:00, 51.23it/s]

Generated data n=19860

train batches: 100%| | 311/311 [06:56<00:00, 1.34s/it, loss: 24.524]

Adjusting learning rate of group 0 to 4.8125e-08.
Average epoch loss: 15.880071403895927
Saved model to: /Users/johnsonzhou/git/comp90042-project/result/models/model_03a_bert_base_triplet_2023_05_03_22_35.pth

dev batches: 100%| | 32/32 [00:07<00:00, 4.35it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]

Average cos sim (pos, neg): 0.972252, 0.867923
Epoch: 1 of 100

Generate siamese dataset

claims: 100%| | 1228/1228 [00:24<00:00, 50.65it/s]

Generated data n=19860

train batches: 100%| | 311/311 [06:56<00:00, 1.34s/it, loss: 20.762]

Adjusting learning rate of group 0 to 7.6250e-08.
Average epoch loss: 15.330983336715454
Saved model to: /Users/johnsonzhou/git/comp90042-project/result/models/model_03a_bert_base_triplet_2023_05_03_22_35.pth

dev batches: 100%| | 32/32 [00:07<00:00, 4.44it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]

Average cos sim (pos, neg): 0.972252, 0.867923
Epoch: 2 of 100

Generate siamese dataset

claims: 100%| | 1228/1228 [00:25<00:00, 48.80it/s]

Generated data n=19860

train batches: 100%| | 311/311 [06:57<00:00, 1.34s/it, loss: 15.610]

Adjusting learning rate of group 0 to 1.0437e-07.
Average epoch loss: 14.469343896080826
Saved model to: /Users/johnsonzhou/git/comp90042-project/result/models/model_03a_bert_base_triplet_2023_05_03_22_35.pth

dev batches: 100%| | 32/32 [00:07<00:00, 4.37it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]

Average cos sim (pos, neg): 0.972252, 0.867923
Epoch: 3 of 100

Generate siamese dataset

claims: 100%| | 1228/1228 [00:24<00:00, 50.51it/s]

Generated data n=19860

train batches: 100%| | 311/311 [06:58<00:00, 1.34s/it, loss: 15.009]

Adjusting learning rate of group 0 to 1.3250e-07.
Average epoch loss: 12.816365626847245
Saved model to: /Users/johnsonzhou/git/comp90042-project/result/models/model_03a_bert_base_triplet_2023_05_03_22_35.pth
dev batches: 100%| | 32/32 [00:07<00:00, 4.35it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]
Average cos sim (pos, neg): 0.972252, 0.867923
Epoch: 4 of 100

Generate siamese dataset

claims: 100%| | 1228/1228 [00:24<00:00, 49.62it/s]

Generated data n=19860

train batches: 100%| | 311/311 [06:55<00:00, 1.33s/it, loss: 7.128]

Adjusting learning rate of group 0 to 1.6062e-07.
Average epoch loss: 10.058780984479897
Saved model to: /Users/johnsonzhou/git/comp90042-project/result/models/model_03a_bert_base_triplet_2023_05_03_22_35.pth
dev batches: 100%| | 32/32 [00:07<00:00, 4.37it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]
Average cos sim (pos, neg): 0.972252, 0.867923
Epoch: 5 of 100

Generate siamese dataset

claims: 100%| | 1228/1228 [00:24<00:00, 50.23it/s]

Generated data n=19860

train batches: 100%| | 311/311 [06:54<00:00, 1.33s/it, loss: 9.192]

Adjusting learning rate of group 0 to 1.8875e-07.
Average epoch loss: 7.877253962866363
Saved model to: /Users/johnsonzhou/git/comp90042-project/result/models/model_03a_bert_base_triplet_2023_05_03_22_35.pth
dev batches: 100%| | 32/32 [00:07<00:00, 4.37it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]
Average cos sim (pos, neg): 0.972252, 0.867923
Epoch: 6 of 100

Generate siamese dataset

claims: 100%| | 1228/1228 [00:24<00:00, 49.17it/s]

Generated data n=19860

train batches: 100%| | 311/311 [06:55<00:00, 1.33s/it, loss: 9.387]

Adjusting learning rate of group 0 to 1.8875e-07.
Average epoch loss: 6.582534527088669
Saved model to: /Users/johnsonzhou/git/comp90042-project/result/models/model_03a_bert_base_triplet_2023_05_03_22_35.pth

dev batches: 100%| | 32/32 [00:07<00:00, 4.38it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]

Average cos sim (pos, neg): 0.972252, 0.867923
Epoch: 7 of 100

Generate siamese dataset

claims: 100%| | 1228/1228 [00:24<00:00, 49.98it/s]

Generated data n=19860

train batches: 100%| | 311/311 [06:54<00:00, 1.33s/it, loss: 7.302]

Adjusting learning rate of group 0 to 1.8875e-07.

Average epoch loss: 5.956192210364572

Saved model to: /Users/johnsonzhou/git/comp90042-project/result/models/model_03a_bert_base_triplet_2023_05_03_22_35.pth

dev batches: 100%| | 32/32 [00:07<00:00, 4.33it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]

Average cos sim (pos, neg): 0.972252, 0.867923
Epoch: 8 of 100

Generate siamese dataset

claims: 100%| | 1228/1228 [00:25<00:00, 47.91it/s]

Generated data n=19860

train batches: 100%| | 311/311 [06:53<00:00, 1.33s/it, loss: 8.414]

Adjusting learning rate of group 0 to 1.8875e-07.

Average epoch loss: 5.492325866720684

Saved model to: /Users/johnsonzhou/git/comp90042-project/result/models/model_03a_bert_base_triplet_2023_05_03_22_35.pth

dev batches: 100%| | 32/32 [00:07<00:00, 4.40it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]

Average cos sim (pos, neg): 0.972252, 0.867923
Epoch: 9 of 100

Generate siamese dataset

claims: 100%| | 1228/1228 [00:24<00:00, 50.85it/s]

Generated data n=19860

train batches: 100%| | 311/311 [06:51<00:00, 1.32s/it, loss: 8.189]

```

Adjusting learning rate of group 0 to 1.8875e-07.
Average epoch loss: 5.266779184916395
Saved model to: /Users/johnsonzhou/git/comp90042-project/result/models/model_03a
_bert_base_triplet_2023_05_03_22_35.pth

dev batches: 100%|          | 32/32 [00:07<00:00, 4.34it/s, pos cos_sim: 0.972
neg cos_sim: 0.868]

Average cos sim (pos, neg): 0.972252, 0.867923
Epoch: 10 of 100

Generate siamese dataset

claims: 100%|          | 1228/1228 [00:25<00:00, 47.49it/s]

Generated data n=19860

train batches: 8%|          | 25/311 [00:34<06:30, 1.36s/it, loss: 5.156]

```

```

-----
KeyboardInterrupt                                Traceback (most recent call last)
Cell In[36], line 45
      42 optimizer.zero_grad()
      44 # Forward + loss
----> 45 anchor_emb, pos_emb, neg_emb = model(anchor, positive, negative)
      46 loss = loss_fn(anchor=anchor_emb, positive=pos_emb, negative=neg_emb)
      48 # Backward + optimiser

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8
site-packages/torch/nn/modules/module.py:1501, in Module._call_impl(self,
*args, **kwargs)
    1496 # If we don't have any hooks, we want to skip the rest of the logic in
    1497 # this function, and just call forward.
    1498 if not (self._backward_hooks or self._backward_pre_hooks or self.
    _forward_hooks or self._forward_pre_hooks
    1499         or _global_backward_pre_hooks or _global_backward_hooks
    1500         or _global_forward_hooks or _global_forward_pre_hooks):
-> 1501     return forward_call(*args, **kwargs)
    1502 # Do not call functions when jit is used
    1503 full_backward_hooks, non_full_backward_hooks = [], []

Cell In[28], line 33, in SiameseTripletEmbedderBert.forward(self, anchor_texts,
pos_texts, neg_texts)
      31 pos_x = self.tokenizer(pos_texts, **t_kwargs)
      32 if neg_texts:
----> 33     neg_x = self.tokenizer(neg_texts, **t_kwargs)
      35 anchor_x = anchor_x["input_ids"].to(device=self.device)
      36 pos_x = pos_x["input_ids"].to(device=self.device)

```



```

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8
↳site-packages/transformers/tokenization_utils_base.py:2488, in
↳PreTrainedTokenizerBase.__call__(self, text, text_pair, text_target,
↳text_pair_target, add_special_tokens, padding, truncation, max_length, stride
↳is_split_into_words, pad_to_multiple_of, return_tensors,
↳return_token_type_ids, return_attention_mask, return_overflowing_tokens,
↳return_special_tokens_mask, return_offsets_mapping, return_length, verbose,
↳**kwargs)
    2486     if not self._in_target_context_manager:
    2487         self._switch_to_input_mode()
-> 2488     encodings = self._call_one(text=text, text_pair=text_pair,
↳**all_kwargs)
    2489     if text_target is not None:
    2490         self._switch_to_target_mode()

```

```

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8
↳site-packages/transformers/tokenization_utils_base.py:2574, in
↳PreTrainedTokenizerBase._call_one(self, text, text_pair, add_special_tokens,
↳padding, truncation, max_length, stride, is_split_into_words,
↳pad_to_multiple_of, return_tensors, return_token_type_ids,
↳return_attention_mask, return_overflowing_tokens, return_special_tokens_mask,
↳return_offsets_mapping, return_length, verbose, **kwargs)
    2569         raise ValueError(
    2570             f"batch length of `text`: {len(text)} does not match batch
↳length of `text_pair`:"
    2571             f" {len(text_pair)})."
    2572         )
    2573     batch_text_or_text_pairs = list(zip(text, text_pair)) if text_pair
↳is not None else text
-> 2574     return self.batch_encode_plus(
    2575         batch_text_or_text_pairs=batch_text_or_text_pairs,
    2576         add_special_tokens=add_special_tokens,
    2577         padding=padding,
    2578         truncation=truncation,
    2579         max_length=max_length,
    2580         stride=stride,
    2581         is_split_into_words=is_split_into_words,
    2582         pad_to_multiple_of=pad_to_multiple_of,
    2583         return_tensors=return_tensors,
    2584         return_token_type_ids=return_token_type_ids,
    2585         return_attention_mask=return_attention_mask,
    2586         return_overflowing_tokens=return_overflowing_tokens,
    2587         return_special_tokens_mask=return_special_tokens_mask,
    2588         return_offsets_mapping=return_offsets_mapping,
    2589         return_length=return_length,
    2590         verbose=verbose,
    2591         **kwargs,
    2592     )
    2593 else:
    2594     return self.encode_plus(
    2595         text=text,

```

```

2596         text_pair=text_pair,
(...)
2612         **kwargs,
2613     )

```

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8

```

↪site-packages/transformers/tokenization_utils_base.py:2765, in
↪PreTrainedTokenizerBase.batch_encode_plus(self, batch_text_or_text_pairs,
↪add_special_tokens, padding, truncation, max_length, stride,
↪is_split_into_words, pad_to_multiple_of, return_tensors,
↪return_token_type_ids, return_attention_mask, return_overflowing_tokens,
↪return_special_tokens_mask, return_offsets_mapping, return_length, verbose,
↪**kwargs)
    2755 # Backward compatibility for 'truncation_strategy', 'pad_to_max_length'
    2756 padding_strategy, truncation_strategy, max_length, kwargs = self.
↪_get_padding_truncation_strategies(
    2757     padding=padding,
    2758     truncation=truncation,
    (...)
    2762     **kwargs,
    2763 )
-> 2765 return self._batch_encode_plus(
    2766     batch_text_or_text_pairs=batch_text_or_text_pairs,
    2767     add_special_tokens=add_special_tokens,
    2768     padding_strategy=padding_strategy,
    2769     truncation_strategy=truncation_strategy,
    2770     max_length=max_length,
    2771     stride=stride,
    2772     is_split_into_words=is_split_into_words,
    2773     pad_to_multiple_of=pad_to_multiple_of,
    2774     return_tensors=return_tensors,
    2775     return_token_type_ids=return_token_type_ids,
    2776     return_attention_mask=return_attention_mask,
    2777     return_overflowing_tokens=return_overflowing_tokens,
    2778     return_special_tokens_mask=return_special_tokens_mask,
    2779     return_offsets_mapping=return_offsets_mapping,
    2780     return_length=return_length,
    2781     verbose=verbose,
    2782     **kwargs,
    2783 )

```

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8

```

↪site-packages/transformers/tokenization_utils.py:733, in PreTrainedTokenizer.
↪_batch_encode_plus(self, batch_text_or_text_pairs, add_special_tokens,
↪padding_strategy, truncation_strategy, max_length, stride,
↪is_split_into_words, pad_to_multiple_of, return_tensors,
↪return_token_type_ids, return_attention_mask, return_overflowing_tokens,
↪return_special_tokens_mask, return_offsets_mapping, return_length, verbose,
↪**kwargs)
    730 else:
    731     ids, pair_ids = ids_or_pair_ids

```

```

--> 733 first_ids = get_input_ids(ids)
      734 second_ids = get_input_ids(pair_ids) if pair_ids is not None else None
      735 input_ids.append((first_ids, second_ids))

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8
↳site-packages/transformers/tokenization_utils.py:700, in PreTrainedTokenizer.
↳_batch_encode_plus.<locals>.get_input_ids(text)
      698 def get_input_ids(text):
      699     if isinstance(text, str):
--> 700         tokens = self.tokenize(text, **kwargs)
      701         return self.convert_tokens_to_ids(tokens)
      702     elif isinstance(text, (list, tuple)) and len(text) > 0 and
↳isinstance(text[0], str):

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8
↳site-packages/transformers/tokenization_utils.py:547, in PreTrainedTokenizer.
↳tokenize(self, text, **kwargs)
      545         tokenized_text.append(token)
      546     else:
--> 547         tokenized_text.extend(self._tokenize(token))
      548 # ["This", " is", " something", "<special_token_1>", "else"]
      549 return tokenized_text

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8
↳site-packages/transformers/models/bert/tokenization_bert.py:244, in
↳BertTokenizer._tokenize(self, text)
      242 split_tokens = []
      243 if self.do_basic_tokenize:
--> 244     for token in self.basic_tokenizer.tokenize(text, never_split=self.
↳all_special_tokens):
      245
      246         # If the token is part of the never_split set
      247         if token in self.basic_tokenizer.never_split:
      248             split_tokens.append(token)

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8
↳site-packages/transformers/models/bert/tokenization_bert.py:431, in
↳BasicTokenizer.tokenize(self, text, never_split)
      429         elif self.strip_accents:
      430             token = self._run_strip_accents(token)
--> 431         split_tokens.extend(self._run_split_on_punc(token, never_split))
      433 output_tokens = whitespace_tokenize(" ".join(split_tokens))
      434 return output_tokens

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8
↳site-packages/transformers/models/bert/tokenization_bert.py:457, in
↳BasicTokenizer._run_split_on_punc(self, text, never_split)
      455 while i < len(chars):
      456     char = chars[i]

```

```

--> 457     if _is_punctuation(char):
      458         output.append([char])
      459         start_new_word = True

```

```

File /opt/homebrew/Caskroom/miniconda/base/envs/comp90042_project/lib/python3.8
↳site-packages/transformers/tokenization_utils.py:299, in _is_punctuation(char)
      294 cp = ord(char)
      295 # We treat all non-letter/number ASCII as punctuation.
      296 # Characters such as "^", "$", and "`" are not in the Unicode
      297 # Punctuation class but we treat them as punctuation anyways, for
      298 # consistency.
--> 299 if (cp >= 33 and cp <= 47) or (cp >= 58 and cp <= 64) or (cp >= 91 and
↳cp <= 96) or (cp >= 123 and cp <= 126):
      300     return True
      301 cat = unicodedata.category(char)

```

KeyboardInterrupt:

```

[ ]: @dataclass
      class SentenceSimilarity:
          label:int
          score:float
          claim_text:str
          evidence_text:str

```

```

[ ]: # Run evaluation before training to establish baseline
      model.eval()

      dev_batches = tqdm(dev_dataloader, desc="dev batches")
      epoch_pos_cos_sim = []
      epoch_neg_cos_sim = []
      sentence_rankings = []
      for batch in dev_batches:
          claim_texts, evidence_texts, labels = batch

          # Forward
          claim_emb, evidence_emb = model(claim_texts, evidence_texts)

          # Cosine similarity
          cos_sim = torch.cosine_similarity(claim_emb, evidence_emb)
          labelled_cos_sim = cos_sim * labels
          pos_cos_sim = labelled_cos_sim[torch.where(labelled_cos_sim > 0)]
          neg_cos_sim = labelled_cos_sim[torch.where(labelled_cos_sim < 0)]

          batch_pos_cos_sim = torch.mean(pos_cos_sim).cpu().item()
          batch_neg_cos_sim = torch.mean(neg_cos_sim).cpu().item() * -1

```

```

epoch_pos_cos_sim.append(batch_pos_cos_sim)
epoch_neg_cos_sim.append(batch_neg_cos_sim)

for claim_text, evidence_text, label, score in \
    zip(
        claim_texts, evidence_texts,
        labels.detach().cpu().numpy(), cos_sim.detach().cpu().numpy()
    ):
    sentence_rankings.append(SentenceSimilarity(
        label=label,
        score=score,
        claim_text=claim_text,
        evidence_text=evidence_text
    ))

dev_batches.postfix = f"pos cos_sim: {batch_pos_cos_sim:.3f}" + \
    f" neg cos_sim: {batch_neg_cos_sim:.3f}"

continue

```

```

dev batches: 100%|          | 32/32 [00:08<00:00, 3.84it/s, pos cos_sim: 0.941
neg cos_sim: 0.905]

```

```
[ ]: sentence_rankings[:50]
```

```

[ ]: [SentenceSimilarity(label=1, score=0.98813456, claim_text='[South Australia] has
the most expensive electricity in the world.', evidence_text='[citation needed]
South Australia has the highest retail price for electricity in the country.'),
SentenceSimilarity(label=1, score=0.98029995, claim_text='[South Australia] has
the most expensive electricity in the world.', evidence_text='"South Australia
has the highest power prices in the world."'),
SentenceSimilarity(label=-1, score=0.91536856, claim_text='[South Australia]
has the most expensive electricity in the world.', evidence_text='"How the
oceans absorb carbon dioxide is critical for predicting climate change."'),
SentenceSimilarity(label=-1, score=0.8487079, claim_text='[South Australia] has
the most expensive electricity in the world.', evidence_text='Families are
forced into increasing poverty, some facing a daily struggle to pay their rent
and put food on their table.'),
SentenceSimilarity(label=-1, score=0.9821, claim_text='[South Australia] has
the most expensive electricity in the world.', evidence_text='They described the
intermediate scenario as the most likely, and that real-world greenhouse gas
forcing had been closest to this scenario.'),
SentenceSimilarity(label=-1, score=0.97179466, claim_text='[South Australia]
has the most expensive electricity in the world.', evidence_text='Absorption of
infrared light at the vibrational frequencies of atmospheric carbon dioxide
traps energy near the surface, warming the surface and the lower atmosphere.'),
SentenceSimilarity(label=-1, score=0.97337914, claim_text='[South Australia]

```

has the most expensive electricity in the world.', evidence_text='Scientists Reach 100% Consensus on Anthropogenic Global Warming.'),

SentenceSimilarity(label=-1, score=0.91890705, claim_text='[South Australia] has the most expensive electricity in the world.', evidence_text='The Labour Party proposed a renegotiation of the withdrawal agreement (towards a closer post-withdrawal relationship with the EU) and would then put this forward as an option in a referendum alongside the option of remaining in the EU.'),

SentenceSimilarity(label=-1, score=0.97250104, claim_text='[South Australia] has the most expensive electricity in the world.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.98607576, claim_text='[South Australia] has the most expensive electricity in the world.', evidence_text='The 2007 melt season let to a minimum 39% below the 1979-2000 average, and for the first time in human memory, the fabled Northwest Passage opened completely.'),

SentenceSimilarity(label=-1, score=0.93309855, claim_text='[South Australia] has the most expensive electricity in the world.', evidence_text='Surface heat and freshwater fluxes create global density gradients that drive the thermohaline circulation part of large-scale ocean circulation.'),

SentenceSimilarity(label=-1, score=0.9768435, claim_text='[South Australia] has the most expensive electricity in the world.', evidence_text='The change was attributed to increasingly cold winters in the Arctic stratosphere at an altitude of approximately 20\xa0km (12\xa0mi), a change associated with global warming in a relationship that is still under investigation.'),

SentenceSimilarity(label=1, score=0.988818, claim_text='when 3 per cent of total annual global emissions of carbon dioxide are from humans and Australia prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions reductio\xadn here will have any effect on global climate.', evidence_text='The 2011 UNEP Green Economy report states that "[a]gricultural operations, excluding land use changes, produce approximately 13 per cent of anthropogenic global GHG emissions.'),

SentenceSimilarity(label=1, score=0.9850658, claim_text='when 3 per cent of total annual global emissions of carbon dioxide are from humans and Australia prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions reductio\xadn here will have any effect on global climate.', evidence_text='With a market share of 30% and (potentially) clean electricity, heat pumps could reduce global CO 2 emissions by 8% annually.'),

SentenceSimilarity(label=1, score=0.92869276, claim_text='when 3 per cent of total annual global emissions of carbon dioxide are from humans and Australia prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions reductio\xadn here will have any effect on global climate.', evidence_text='In the modern era, emissions to the atmosphere from volcanoes are approximately 0.645 billion tonnes of CO 2 per year, whereas humans contribute 29 billion tonnes of CO 2 each year.'),

SentenceSimilarity(label=1, score=0.99298954, claim_text='when 3 per cent of total annual global emissions of carbon dioxide are from humans and Australia

prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions
reductio\xadn here will have any effect on global climate.',
evidence_text='Cumulative anthropogenic (i.e., human-emitted) emissions of CO₂
from fossil fuel use are a major cause of global warming, and give some
indication of which countries have contributed most to human-induced climate
change.')

SentenceSimilarity(label=1, score=0.9899789, claim_text='when 3 per cent of
total annual global emissions of carbon dioxide are from humans and Australia
prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions
reductio\xadn here will have any effect on global climate.',
evidence_text='Other countries with fast growing emissions are South Korea,
Iran, and Australia (which apart from the oil rich Persian Gulf states, now has
the highest percapita emission rate in the world).'),

SentenceSimilarity(label=-1, score=0.891354, claim_text='when 3 per cent of
total annual global emissions of carbon dioxide are from humans and Australia
prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions
reductio\xadn here will have any effect on global climate.', evidence_text='"How
the oceans absorb carbon dioxide is critical for predicting climate change".'),

SentenceSimilarity(label=-1, score=0.8602732, claim_text='when 3 per cent of
total annual global emissions of carbon dioxide are from humans and Australia
prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions
reductio\xadn here will have any effect on global climate.',
evidence_text='Families are forced into increasing poverty, some facing a daily
struggle to pay their rent and put food on their table.')

SentenceSimilarity(label=-1, score=0.9704923, claim_text='when 3 per cent of
total annual global emissions of carbon dioxide are from humans and Australia
prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions
reductio\xadn here will have any effect on global climate.', evidence_text='They
described the intermediate scenario as the most likely, and that real-world
greenhouse gas forcing had been closest to this scenario.')

SentenceSimilarity(label=-1, score=0.96001637, claim_text='when 3 per cent of
total annual global emissions of carbon dioxide are from humans and Australia
prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions
reductio\xadn here will have any effect on global climate.',
evidence_text='Absorption of infrared light at the vibrational frequencies of
atmospheric carbon dioxide traps energy near the surface, warming the surface
and the lower atmosphere.')

SentenceSimilarity(label=-1, score=0.95457125, claim_text='when 3 per cent of
total annual global emissions of carbon dioxide are from humans and Australia
prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions
reductio\xadn here will have any effect on global climate.',
evidence_text='Scientists Reach 100% Consensus on Anthropogenic Global
Warming.')

SentenceSimilarity(label=-1, score=0.94295055, claim_text='when 3 per cent of
total annual global emissions of carbon dioxide are from humans and Australia
prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions
reductio\xadn here will have any effect on global climate.', evidence_text='The

Labour Party proposed a renegotiation of the withdrawal agreement (towards a closer post-withdrawal relationship with the EU) and would then put this forward as an option in a referendum alongside the option of remaining in the EU.'),

SentenceSimilarity(label=-1, score=0.9611641, claim_text='when 3 per cent of total annual global emissions of carbon dioxide are from humans and Australia prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions reductio\xadn here will have any effect on global climate.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9823944, claim_text='when 3 per cent of total annual global emissions of carbon dioxide are from humans and Australia prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions reductio\xadn here will have any effect on global climate.', evidence_text='The 2007 melt season let to a minimum 39% below the 1979-2000 average, and for the first time in human memory, the fabled Northwest Passage opened completely.'),

SentenceSimilarity(label=-1, score=0.9148158, claim_text='when 3 per cent of total annual global emissions of carbon dioxide are from humans and Australia prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions reductio\xadn here will have any effect on global climate.', evidence_text='Surface heat and freshwater fluxes create global density gradients that drive the thermohaline circulation part of large-scale ocean circulation.'),

SentenceSimilarity(label=-1, score=0.96838117, claim_text='when 3 per cent of total annual global emissions of carbon dioxide are from humans and Australia prod\xaduces 1.3 per cent of this 3 per cent, then no amount of emissions reductio\xadn here will have any effect on global climate.', evidence_text='The change was attributed to increasingly cold winters in the Arctic stratosphere at an altitude of approximately 20\xad0km (12\xad0mi), a change associated with global warming in a relationship that is still under investigation.'),

SentenceSimilarity(label=1, score=0.9803322, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='Multiple independently produced instrumental datasets confirm that the 2009-2018 decade was 0.93 ± 0.07 \xa0C warmer than the pre-industrial baseline (1850-1900).'),

SentenceSimilarity(label=1, score=0.99900985, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='The planet is now 0.8 \xa0C warmer than in pre-industrial times.'),

SentenceSimilarity(label=-1, score=0.98900735, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='"How the oceans absorb carbon dioxide is critical for predicting climate change".'),

SentenceSimilarity(label=-1, score=0.62686336, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='Families are forced into increasing poverty, some facing a daily struggle to pay their rent and put food on their table.'),

SentenceSimilarity(label=-1, score=0.9486822, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='They

described the intermediate scenario as the most likely, and that real-world greenhouse gas forcing had been closest to this scenario.')

SentenceSimilarity(label=-1, score=0.9591507, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='Absorption of infrared light at the vibrational frequencies of atmospheric carbon dioxide traps energy near the surface, warming the surface and the lower atmosphere.')

SentenceSimilarity(label=-1, score=0.9580598, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='Scientists Reach 100% Consensus on Anthropogenic Global Warming.')

SentenceSimilarity(label=-1, score=0.74508226, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='The Labour Party proposed a renegotiation of the withdrawal agreement (towards a closer post-withdrawal relationship with the EU) and would then put this forward as an option in a referendum alongside the option of remaining in the EU.')

SentenceSimilarity(label=-1, score=0.96165293, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.')

SentenceSimilarity(label=-1, score=0.8916185, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='The 2007 melt season let to a minimum 39% below the 1979-2000 average, and for the first time in human memory, the fabled Northwest Passage opened completely.')

SentenceSimilarity(label=-1, score=0.9897535, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='Surface heat and freshwater fluxes create global density gradients that drive the thermohaline circulation part of large-scale ocean circulation.')

SentenceSimilarity(label=-1, score=0.95222604, claim_text='This means that the world is now 1C warmer than it was in pre-industrial times', evidence_text='The change was attributed to increasingly cold winters in the Arctic stratosphere at an altitude of approximately 20\xa0km (12\xa0mi), a change associated with global warming in a relationship that is still under investigation.')

SentenceSimilarity(label=1, score=0.8877557, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='Genetic disorders are the result of deleterious mutations and can be due to spontaneous mutation in the affected individual, or can be inherited.')

SentenceSimilarity(label=1, score=0.9565435, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='These errors, called mutations, can affect the phenotype of an organism, especially if they occur within the protein coding sequence of a gene.')

SentenceSimilarity(label=1, score=0.9854546, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.',

evidence_text='This is especially useful studying diseases in adults by allowing expression after a certain period of growth, thus eliminating the deleterious effect of gene expression seen during stages of development in model organisms. '),

SentenceSimilarity(label=1, score=0.9894745, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='In addition, research was needed to determine the mechanism by which Zika produced these effects. '),

SentenceSimilarity(label=1, score=0.9654056, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='Zika also appears to have an equal tropism for cells of the developing eye, leading to high rates of eye abnormalities as well. '),

SentenceSimilarity(label=-1, score=0.8974488, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='"How the oceans absorb carbon dioxide is critical for predicting climate change". '),

SentenceSimilarity(label=-1, score=0.8641439, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='Families are forced into increasing poverty, some facing a daily struggle to pay their rent and put food on their table. '),

SentenceSimilarity(label=-1, score=0.97568, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='They described the intermediate scenario as the most likely, and that real-world greenhouse gas forcing had been closest to this scenario. '),

SentenceSimilarity(label=-1, score=0.965821, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='Absorption of infrared light at the vibrational frequencies of atmospheric carbon dioxide traps energy near the surface, warming the surface and the lower atmosphere. '),

SentenceSimilarity(label=-1, score=0.9611697, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='Scientists Reach 100% Consensus on Anthropogenic Global Warming. '),

SentenceSimilarity(label=-1, score=0.93367445, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='The Labour Party proposed a renegotiation of the withdrawal agreement (towards a closer post-withdrawal relationship with the EU) and would then put this forward as an option in a referendum alongside the option of remaining in the EU. ')]

```
[ ]: pos_sentence_rankings = sorted([
    s for s in sentence_rankings
    if s.label == 1
], key=lambda s: s.score)

neg_sentence_rankings = sorted([
    s for s in sentence_rankings
```

```

        if s.label == -1
    ], key=lambda s: s.score, reverse=True)

```

```

[ ]: pos_sentence_rankings[:50]

```

```

[ ]: [SentenceSimilarity(label=1, score=0.66806406, claim_text='More importantly, the
OISM list only contains 39 scientists who specialise in climate science.',
evidence_text='A botanist, plant scientist or phytologist is a scientist who
specialises in this field.'),
SentenceSimilarity(label=1, score=0.6845525, claim_text='Newt Gingrich "teamed
with Nancy Pelosi and Al Gore on global warming."', evidence_text='He was known
as one of the Atari Democrats, later called the "Democrats\' Greens, politicians
who see issues like clean air, clean water and global warming as the key to
future victories for their party."'),
SentenceSimilarity(label=1, score=0.70060915, claim_text='The unlikely
scenarios are now, all of a sudden, becoming more probable than they once were
thought to be,' says Sweet."', evidence_text='The reverse possibility, that a
small group broke off and wandered from India into Western Asia is readily
dismissed as an improbably long migration, again without the least bit of
evidence.'),
SentenceSimilarity(label=1, score=0.7659733, claim_text='More importantly, the
OISM list only contains 39 scientists who specialise in climate science.',
evidence_text='The list includes scientists from several specialities or
disciplines.'),
SentenceSimilarity(label=1, score=0.7696998, claim_text='[...] Constant 24-7
media coverage of every significant storm worldwide just makes it seem that
way."', evidence_text='The severe and widespread damage the storm caused in the
United States, as well as its unusual merger with a frontal system, resulted in
the nicknaming of the hurricane "Superstorm Sandy" by the media, public
officials, and several organizations, including U.S. government agencies.'),
SentenceSimilarity(label=1, score=0.77562284, claim_text='a study that totally
debunks the whole concept of man-made Global Warming', evidence_text='"Warming
\'very likely\' human-made.'),
SentenceSimilarity(label=1, score=0.78157717, claim_text='Satellite
measurements of infrared spectra over the past 40 years observe less energy
escaping to space at the wavelengths associated with CO2.', evidence_text='The
three channels use the same frequency but different carbon dioxide cell
pressure, the corresponding weighting functions peaks at 29\xa0km for channel 1,
37\xa0km for channel 2 and 45\xa0km for channel 3.'),
SentenceSimilarity(label=1, score=0.7819738, claim_text='"Which is to say that
these beans will be eaten by cows, and the cows will convert the beans to meat,
and the humans will eat the meat.', evidence_text='are vehemently against the
consumption of meat and eggs (though they do consume and encourage the
consumption of milk, butter and cheese).'),
SentenceSimilarity(label=1, score=0.7853484, claim_text='Global Warming history
completely coincides with the history of artificial satellites and the use of
microwave frequencies from outer space.', evidence_text='A communications

```

satellite is an artificial satellite that relays and amplifies radio telecommunications signals via a transponder; it creates a communication channel between a source transmitter and a receiver at different locations on Earth. '),

SentenceSimilarity(label=1, score=0.78728116, claim_text="Without the forests' humidity, previously moisture-laden winds blew dry.", evidence_text='Laurel forest and other cloud forest are subtropical and mild temperate forest, found in areas with high humidity and relatively stable and mild temperatures. '),

SentenceSimilarity(label=1, score=0.7923066, claim_text='many scientists were surprised when other researchers subsequently found that ringed and bearded seals (the primary prey of polar bears) north of the Bering Strait especially thrived with a longer open-water season, which is particularly conducive to fishing', evidence_text='The polar bear is the most carnivorous species of bear, and its diet primarily consists of ringed (Pusa hispida) and bearded (Erignathus barbatus) seals. '),

SentenceSimilarity(label=1, score=0.7928746, claim_text='"Austria is today seeing its earliest snowfall in history with 30 to 40 centimetres already predicted in the mountains.', evidence_text="This law of neutrality, passed in late October 1955 (and not the State Treaty itself), ensured that modern Austria would align with neither NATO nor the Soviet bloc, and is considered one of the reasons for Austria's delayed entry into the European Union in 1995."),

SentenceSimilarity(label=1, score=0.8038218, claim_text='Any reasonable person can recognize both positives and negatives among the policy proposals of both Tories and Labour.', evidence_text='In a speech in Tynemouth the next day, May said Labour had "deserted" working-class voters, criticised Labour\'s policy proposals and said Britain\'s future depended on making a success of Brexit. '),

SentenceSimilarity(label=1, score=0.8052311, claim_text='The gas builds up in the soil, forming mounds called 'pingoes.', evidence_text='At the next larger scale, soil structures called peds or more commonly soil aggregates are created from the soil separates when iron oxides, carbonates, clay, silica and humus, coat particles and cause them to adhere into larger, relatively stable secondary structures. '),

SentenceSimilarity(label=1, score=0.8104248, claim_text='many scientists were surprised when other researchers subsequently found that ringed and bearded seals (the primary prey of polar bears) north of the Bering Strait especially thrived with a longer open-water season, which is particularly conducive to fishing', evidence_text='The polar bear is the most carnivorous member of the bear family, and throughout most of its range, its diet primarily consists of ringed (Pusa hispida) and bearded seals (Erignathus barbatus). '),

SentenceSimilarity(label=1, score=0.81556827, claim_text='Satellite measurements of infrared spectra over the past 40 years observe less energy escaping to space at the wavelengths associated with CO2.', evidence_text='They measure radiances in various wavelength bands. '),

SentenceSimilarity(label=1, score=0.8236294, claim_text='Protecting and restoring forests would reduce 18% of emissions by 2030', evidence_text='Japan needs to reduce 26% of green house gas emission from 2013 by 2030 to accomplish Paris Agreement and is trying to reduce 2% of them by forestry. '),

SentenceSimilarity(label=1, score=0.82607484, claim_text='"Another global

warming myth comes crashing down.", evidence_text='Delingpole has engaged in climate change denialism; in 2009 he wrote of "The conspiracy behind the Anthropogenic Global Warming myth".'),

SentenceSimilarity(label=1, score=0.82813877, claim_text='"Which is to say that these beans will be eaten by cows, and the cows will convert the beans to meat, and the humans will eat the meat.', evidence_text='To obtain milk from dairy cattle, cows are made pregnant to induce lactation; they are kept lactating for three to seven years, then slaughtered.'),

SentenceSimilarity(label=1, score=0.82939434, claim_text='Any reasonable person can recognize both positives and negatives among the policy proposals of both Tories and Labour.', evidence_text='The Liberal Democrats, the Greens, the SNP and Labour all support a ban on fracking, whilst the Conservatives propose approving fracking on a case-by-case basis.'),

SentenceSimilarity(label=1, score=0.82969666, claim_text='\''To suddenly label CO2 as a "pollutant" is a disservice to a gas that has played an enormous role in the development and sustainability of all life on this wonderful Earth.', evidence_text='Emissions trading (also known as cap and trade) is a market-based approach to controlling pollution by providing economic incentives for achieving reductions in the emissions of pollutants.'),

SentenceSimilarity(label=1, score=0.832629, claim_text='The heatwave we now have in Europe is not something that was expected with just 1C of warming', evidence_text='Ruse, Bulgaria hit 36.6\xa0°C (97.9\xa0°F) on the 13th making it the warmest spot in Europe.'),

SentenceSimilarity(label=1, score=0.8351672, claim_text='In many other cases, though - hurricanes, for example - the linkage to global warming for particular trends is uncertain or disputed.', evidence_text='Although there are a few areas of linkage, the relationship between the two is weak.'),

SentenceSimilarity(label=1, score=0.83735085, claim_text='The unlikely scenarios are now, all of a sudden, becoming more probable than they once were thought to be,' says Sweet."', evidence_text='"In recent decades, a few specialists have continued to support this interpretation, and Peter Schrijver has said that 'to a large extent, it is linguistics that is responsible for thinking in terms of drastic scenarios' about demographic change in late Roman Britain."'),

SentenceSimilarity(label=1, score=0.83778113, claim_text='But [climate scientists] say that aspects of the case of Hurricane Harvey-and the recent history of tropical cyclones worldwide-suggest global warming is making a bad situation worse.', evidence_text='Global warming not only causes changes in tropical cyclones, it may also make some impacts from them worse via sea level rise.'),

SentenceSimilarity(label=1, score=0.84041834, claim_text='Many of the world's coral reefs are already barren or in a state of constant decline.', evidence_text='Aquaculture is showing promise as a potentially effective tool for restoring coral reefs, which have been declining around the world.'),

SentenceSimilarity(label=1, score=0.84201324, claim_text='Heat is continuing to build up in the subsurface ocean.', evidence_text='It plays an important role in supplying heat to the polar regions, and thus in sea ice regulation.'),

SentenceSimilarity(label=1, score=0.84264547, claim_text='The grasslands, crops, forests and territorial waters of Australia absorb more carbon dioxide than Australia emits.', evidence_text='They absorb a huge amount of carbon dioxide, combating climate change.'),

SentenceSimilarity(label=1, score=0.84488255, claim_text='"Which is to say that these beans will be eaten by cows, and the cows will convert the beans to meat, and the humans will eat the meat.', evidence_text='conversion of human-inedible residues of food crops.'),

SentenceSimilarity(label=1, score=0.8462859, claim_text='Many of the world's coral reefs are already barren or in a state of constant decline.', evidence_text='For example, Midway Atoll in Hawaii supports nearly three million seabirds, including two-thirds (1.5 million) of the global population of Laysan albatross, and one-third of the global population of black-footed albatross.'),

SentenceSimilarity(label=1, score=0.8466495, claim_text='But each serial adjustment has tended to make the early years colder, which increases the warming trend.', evidence_text='Climate change adaptation is "the adjustment in natural or human systems in response to actual or expected climatic stimuli or their effects, which moderates harm or exploits beneficial opportunities".'),

SentenceSimilarity(label=1, score=0.8477817, claim_text='"Which is to say that these beans will be eaten by cows, and the cows will convert the beans to meat, and the humans will eat the meat.', evidence_text='Buffalo are generally fed on coarse feeds; they convert them into the protein-rich lean meat.'),

SentenceSimilarity(label=1, score=0.8496723, claim_text='[Riebesell] is a world authority on the topic and has typically communicated cautiously about the effects of acidification.', evidence_text='Perhaps one of the most recent adverse effects of climate change to be explored is that of ocean acidification.'),

SentenceSimilarity(label=1, score=0.84969604, claim_text='In the past, warming has never been a threat to life on Earth.', evidence_text='Given the potential threat to marine ecosystems and its ensuing impact on human society and economy, especially as it acts in conjunction with anthropogenic global warming, there is an urgent need for immediate action.'),

SentenceSimilarity(label=1, score=0.8514199, claim_text='Newt Gingrich "teamed with Nancy Pelosi and Al Gore on global warming."', evidence_text='Gingrich has identified education as "the number one factor in our future prosperity", and has partnered with Al Sharpton and Education Secretary Arne Duncan on education issues.'),

SentenceSimilarity(label=1, score=0.8535549, claim_text='many scientists were surprised when other researchers subsequently found that ringed and bearded seals (the primary prey of polar bears) north of the Bering Strait especially thrived with a longer open-water season, which is particularly conducive to fishing', evidence_text='Natural predators of the bearded seal include polar bears, who rely on these seals as a major food source.'),

SentenceSimilarity(label=1, score=0.8573083, claim_text='Greenpeace didn't save the whales, switching from whale oil to petroleum and palm oil did', evidence_text='Greenpeace is also critical of extracting petroleum from oil sands and has used direct action to block operations at the Athabasca oil sands

in Canada.'),

SentenceSimilarity(label=1, score=0.8577589, claim_text='Protecting and restoring forests would reduce 18% of emissions by 2030', evidence_text='The UN estimate deforestation and forest degradation to make up 17% of global carbon emissions, which makes it the second most polluting sector, following the energy industry.'),

SentenceSimilarity(label=1, score=0.8635001, claim_text='many scientists were surprised when other researchers subsequently found that ringed and bearded seals (the primary prey of polar bears) north of the Bering Strait especially thrived with a longer open-water season, which is particularly conducive to fishing', evidence_text='Ringed seals are one of the primary prey of polar bears and killer whales, and have long been a component of the diet of indigenous people of the Arctic.'),

SentenceSimilarity(label=1, score=0.86742353, claim_text='The heatwave we now have in Europe is not something that was expected with just 1C of warming', evidence_text='2018's temperature was 16.1°C (61.0°F), meaning it ranks as the 18th warmest June recorded in England in the past 359 years, also being the warmest since 1976.'),

SentenceSimilarity(label=1, score=0.87134427, claim_text='Austria is today seeing its earliest snowfall in history with 30 to 40 centimetres already predicted in the mountains.', evidence_text='Austria lies between latitudes 46° and 49° N, and longitudes 9° and 18° E. It can be divided into five areas, the biggest being the Eastern Alps, which constitute 62% of the nation's total area.'),

SentenceSimilarity(label=1, score=0.8744137, claim_text='With 32 years of rapidly increasing global temperatures and only a minor increase in global CO2 emissions, followed by 33 years of slowly cooling global temperatures with rapid increases in global CO2 emissions, it was deceitful for the IPCC to make any claim that CO2 emissions were primarily responsible for observed 20th century global warming.', evidence_text='The global warming observed over the past 50 years is due primarily to human-induced emissions of heat-trapping gases.'),

SentenceSimilarity(label=1, score=0.8782981, claim_text='Protecting and restoring forests would reduce 18% of emissions by 2030', evidence_text='Targets for the year 2030: Reduce GHG emission by 40% from the level of 1990.'),

SentenceSimilarity(label=1, score=0.87879854, claim_text='In 2007, the Northern Hemisphere reached a record low in ice coverage and the Northwest Passage was opened.', evidence_text='On August 21, 2007, the Northwest Passage became open to ships without the need of an icebreaker.'),

SentenceSimilarity(label=1, score=0.88012344, claim_text='Greg Hunt CSIRO research shows carbon emissions can be reduced by 20 per cent over 40 years using nature, soils and trees.', evidence_text='There is a large potential for future reductions in emissions by a combination of activities, including energy conservation and increased energy efficiency; the use of low-carbon energy technologies, such as renewable energy, nuclear energy, and carbon capture and storage; decarbonizing buildings and transport; and enhancing carbon sinks through, for example, reforestation and preventing deforestation.'),

SentenceSimilarity(label=1, score=0.88481367, claim_text='"Which is to say that these beans will be eaten by cows, and the cows will convert the beans to meat, and the humans will eat the meat.', evidence_text='Carnivores require meat in their diet for survival, but carnists choose to eat meat based on their beliefs.'),

SentenceSimilarity(label=1, score=0.8874148, claim_text='"Each unit of CO2 you put into the atmosphere has less and less of a warming impact.', evidence_text='The Kyoto Protocol is an international treaty which extends the 1992 United Nations Framework Convention on Climate Change (UNFCCC) that commits state parties to reduce greenhouse gas emissions, based on the scientific consensus that (part one) global warming is occurring and (part two) it is extremely likely that human-made CO2 emissions have predominantly caused it.'),

SentenceSimilarity(label=1, score=0.8877557, claim_text='"As it happens, Zika may also be a good model of the second worrying effect - disease mutation.', evidence_text='Genetic disorders are the result of deleterious mutations and can be due to spontaneous mutation in the affected individual, or can be inherited.'),

SentenceSimilarity(label=1, score=0.88840187, claim_text='The grasslands, crops, forests and territorial waters of Australia absorb more carbon dioxide than Australia emits.', evidence_text='It decomposes and turns into carbon dioxide (CO2), which is released into the atmosphere.'),

SentenceSimilarity(label=1, score=0.88853514, claim_text='Concentrated in the atmosphere, these gases do not allow the warmth of the sun's rays reflected by the earth to be dispersed in space.', evidence_text='Ozone acts as a greenhouse gas, absorbing some of the infrared energy emitted by the earth.'])]

```
[ ]: neg_sentence_rankings[:50]
```

```
[ ]: [SentenceSimilarity(label=-1, score=0.99717546, claim_text='The team of climate scientists notes that in failing to predict the warming 'hiatus' in the beginning of the 21st century, the Intergovernmental Panel on Climate Change (IPCC) models overestimated temperature increases...', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),
```

```
    SentenceSimilarity(label=-1, score=0.99610794, claim_text='Once natural influences, in particular the impact of El Niño and La Niña, are removed from the recent temperature record, there is no evidence of a significant change in the human contribution to climate change.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),
```

```
    SentenceSimilarity(label=-1, score=0.9957582, claim_text='Scientists studying Antarctica sea ice warn a rise in accumulation could spark the next ice age.', evidence_text='Surface heat and freshwater fluxes create global density gradients that drive the thermohaline circulation part of large-scale ocean circulation.'),
```


SentenceSimilarity(label=-1, score=0.9952076, claim_text='The IPCC simply updated their temperature history graphs to show the best data available at the time.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9951062, claim_text='Global warming is driving major melting on the surface of Greenland's glaciers and is speeding up their travel into the sea.', evidence_text='How the oceans absorb carbon dioxide is critical for predicting climate change.'),

SentenceSimilarity(label=-1, score=0.99447334, claim_text='[Ice] is expanding in much of Antarctica, contrary to the widespread public belief that global warming is melting the continental ice cap.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9944259, claim_text='Some global warming "skeptics" argue that the Earth's climate sensitivity is so low that a doubling of atmospheric CO2 will result in a surface temperature change on the order of 1°C or less, and that therefore global warming is nothing to worry about.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.99440724, claim_text='Multiple lines of evidence indicate Greenland's ice loss is accelerating and will contribute sea level rise in the order of metres over the next few centuries.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9943454, claim_text='In many other cases, though - hurricanes, for example - the linkage to global warming for particular trends is uncertain or disputed.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.994221, claim_text='While there are isolated cases of growing glaciers, the overwhelming trend in glaciers worldwide is retreat.', evidence_text='Scientists Reach 100% Consensus on Anthropogenic Global Warming.'),

SentenceSimilarity(label=-1, score=0.9941925, claim_text='The most recent IPCC report lays out a future if we limit global heating to 1.5°C instead of the Paris Agreement's 2°C.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.994007, claim_text='By 2050 there's a scientific consensus that we reached the tipping point for ice sheets in Greenland and the West Antarctic', evidence_text='The change was attributed to increasingly cold winters in the Arctic stratosphere at an altitude of approximately 20\000km (12\000mi), a change associated with global warming in a

relationship that is still under investigation.'),

SentenceSimilarity(label=-1, score=0.9938715, claim_text='Many of the world's coral reefs are already barren or in a state of constant decline.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9938112, claim_text='Global Warming history completely coincides with the history of artificial satellites and the use of microwave frequencies from outer space.', evidence_text='The change was attributed to increasingly cold winters in the Arctic stratosphere at an altitude of approximately 20\xa0km (12\xa0mi), a change associated with global warming in a relationship that is still under investigation.'),

SentenceSimilarity(label=-1, score=0.9937221, claim_text='The team of climate scientists notes that in failing to predict the warming 'hiatus' in the beginning of the 21st century, the Intergovernmental Panel on Climate Change (IPCC) models overestimated temperature increases...', evidence_text='The change was attributed to increasingly cold winters in the Arctic stratosphere at an altitude of approximately 20\xa0km (12\xa0mi), a change associated with global warming in a relationship that is still under investigation.'),

SentenceSimilarity(label=-1, score=0.9935299, claim_text='While there are isolated cases of growing glaciers, the overwhelming trend in glaciers worldwide is retreat.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.99346936, claim_text='Global Warming history completely coincides with the history of artificial satellites and the use of microwave frequencies from outer space.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9934211, claim_text='Thick arctic sea ice is in rapid retreat.', evidence_text='Scientists Reach 100% Consensus on Anthropogenic Global Warming.'),

SentenceSimilarity(label=-1, score=0.9934199, claim_text='By 2050 there's a scientific consensus that we reached the tipping point for ice sheets in Greenland and the West Antarctic', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9933, claim_text='Lake-bottom sediments in Florida tell us that recent major hurricane activity in the Gulf of Mexico has been less frequent than in centuries past.', evidence_text='Surface heat and freshwater fluxes create global density gradients that drive the thermohaline circulation part of large-scale ocean circulation.'),

SentenceSimilarity(label=-1, score=0.9931588, claim_text='CO2 limits won't cool the planet.', evidence_text='Scientists Reach 100% Consensus on Anthropogenic Global Warming.'),

SentenceSimilarity(label=-1, score=0.9931034, claim_text='[...] Constant 24-7

media coverage of every significant storm worldwide just makes it seem that way."', evidence_text="How the oceans absorb carbon dioxide is critical for predicting climate change".'),

SentenceSimilarity(label=-1, score=0.9929631, claim_text='Water vapor helps trap heat, and is a far the strongest of the major greenhouse gases, contributing 36-72 percent of the greenhouse effect.', evidence_text='Absorption of infrared light at the vibrational frequencies of atmospheric carbon dioxide traps energy near the surface, warming the surface and the lower atmosphere.'),

SentenceSimilarity(label=-1, score=0.99296236, claim_text="Some global warming 'skeptics' argue that the Earth's climate sensitivity is so low that a doubling of atmospheric CO2 will result in a surface temperature change on the order of 1°C or less, and that therefore global warming is nothing to worry about.", evidence_text='The change was attributed to increasingly cold winters in the Arctic stratosphere at an altitude of approximately 20\xa0km (12\xa0mi), a change associated with global warming in a relationship that is still under investigation.'),

SentenceSimilarity(label=-1, score=0.99291956, claim_text='The most recent IPCC report lays out a future if we limit global heating to 1.5°C instead of the Paris Agreement's 2°C.', evidence_text='Surface heat and freshwater fluxes create global density gradients that drive the thermohaline circulation part of large-scale ocean circulation.'),

SentenceSimilarity(label=-1, score=0.9928293, claim_text='The team of climate scientists notes that in failing to predict the warming 'hiatus' in the beginning of the 21st century, the Intergovernmental Panel on Climate Change (IPCC) models overestimated temperature increases...', evidence_text='Absorption of infrared light at the vibrational frequencies of atmospheric carbon dioxide traps energy near the surface, warming the surface and the lower atmosphere.'),

SentenceSimilarity(label=-1, score=0.99255586, claim_text='The CO2 amplifies the warming and mixes through the atmosphere, spreading warming throughout the planet.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9925486, claim_text='Evidence is building that net cloud feedback is likely positive and unlikely to be strongly negative.', evidence_text="How the oceans absorb carbon dioxide is critical for predicting climate change".'),

SentenceSimilarity(label=-1, score=0.99247974, claim_text='Venus very likely underwent a runaway or 'moist' greenhouse phase earlier in its history, and today is kept hot by a dense CO2 atmosphere.', evidence_text='The change was attributed to increasingly cold winters in the Arctic stratosphere at an altitude of approximately 20\xa0km (12\xa0mi), a change associated with global warming in a relationship that is still under investigation.'),

SentenceSimilarity(label=-1, score=0.9924191, claim_text='The IPCC simply updated their temperature history graphs to show the best data available at the time.', evidence_text='The change was attributed to increasingly cold winters in the Arctic stratosphere at an altitude of approximately 20\xa0km (12\xa0mi), a change associated with global warming in a relationship that is still under

investigation.'),

SentenceSimilarity(label=-1, score=0.9923841, claim_text='Once natural influences, in particular the impact of El Niño and La Niña, are removed from the recent temperature record, there is no evidence of a significant change in the human contribution to climate change.', evidence_text='Absorption of infrared light at the vibrational frequencies of atmospheric carbon dioxide traps energy near the surface, warming the surface and the lower atmosphere.'),

SentenceSimilarity(label=-1, score=0.9922949, claim_text='While there are isolated cases of growing glaciers, the overwhelming trend in glaciers worldwide is retreat.', evidence_text='They described the intermediate scenario as the most likely, and that real-world greenhouse gas forcing had been closest to this scenario.'),

SentenceSimilarity(label=-1, score=0.99215627, claim_text='Global human emissions are only 3 per cent of total annual emissions.', evidence_text='Surface heat and freshwater fluxes create global density gradients that drive the thermohaline circulation part of large-scale ocean circulation.'),

SentenceSimilarity(label=-1, score=0.99215406, claim_text='There are a myriad of other radiative forcings that affect the planet's energy imbalance.', evidence_text='Surface heat and freshwater fluxes create global density gradients that drive the thermohaline circulation part of large-scale ocean circulation.'),

SentenceSimilarity(label=-1, score=0.9920883, claim_text='With more CO2 in the atmosphere, the challenge [feeding 2.5 billion more people] can and will be met.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.99205005, claim_text='This means the jet stream meanders more, with big loops bringing warm air to the frozen north and cold air into warmer southern climes.', evidence_text='How the oceans absorb carbon dioxide is critical for predicting climate change.'),

SentenceSimilarity(label=-1, score=0.99198157, claim_text='CFCs contribute to global warming at a small level.', evidence_text='Scientists Reach 100% Consensus on Anthropogenic Global Warming.'),

SentenceSimilarity(label=-1, score=0.9919663, claim_text='But like most claims regarding global warming, the real effect is small, probably temporary, and most likely due to natural weather patterns', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.99193513, claim_text='Nor is there evidence of an increase in floods globally.', evidence_text='Surface heat and freshwater fluxes create global density gradients that drive the thermohaline circulation part of large-scale ocean circulation.'),

SentenceSimilarity(label=-1, score=0.99191785, claim_text='The noted oceanographer Walter Munk referred to sea-level rise as an "enigma", evidence_text='They described the intermediate scenario as the most likely, and that real-world greenhouse gas forcing had been closest to this scenario.'),

SentenceSimilarity(label=-1, score=0.9918206, claim_text='Through its impacts on the climate, CO2 presents a danger to public health and welfare, and thus qualifies as an air pollutant', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9917978, claim_text='While there are isolated cases of growing glaciers, the overwhelming trend in glaciers worldwide is retreat.', evidence_text='Absorption of infrared light at the vibrational frequencies of atmospheric carbon dioxide traps energy near the surface, warming the surface and the lower atmosphere.'),

SentenceSimilarity(label=-1, score=0.99175924, claim_text='Scientists retracted claim that sea levels are rising.', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9917398, claim_text="'Global warming' and 'climate change' mean different things and have both been used for decades.", evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.9917101, claim_text='IPCC graph showing accelerating trends is misleading', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.99167573, claim_text='IPCC report warning last week the world is "nowhere near on track" to meet its Paris commitments', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.99147606, claim_text='Greenland has only lost a tiny fraction of its ice mass', evidence_text='Projections based on the Special Report on Emissions Scenarios suggest warming over the 21st century at a more rapid rate than that experienced for at least the last 10,000 years.'),

SentenceSimilarity(label=-1, score=0.99142253, claim_text='Once natural influences, in particular the impact of El Niño and La Niña, are removed from the recent temperature record, there is no evidence of a significant change in the human contribution to climate change.', evidence_text='The change was attributed to increasingly cold winters in the Arctic stratosphere at an altitude of approximately 20\xa0km (12\xa0mi), a change associated with global warming in a relationship that is still under investigation.'),

SentenceSimilarity(label=-1, score=0.9914092, claim_text='If there were [carbon emissions], we could not see because most carbon is black.', evidence_text='"How the oceans absorb carbon dioxide is critical for predicting climate change".'),

SentenceSimilarity(label=-1, score=0.99128646, claim_text='Global Warming history completely coincides with the history of artificial satellites and the use of microwave frequencies from outer space.', evidence_text='They described the intermediate scenario as the most likely, and that real-world greenhouse gas forcing had been closest to this scenario.'])