FINETUNING CLIMATEGPT

KAHLIA HOGG

01 Carbon Footprint

02 ClimateGPT

03 Dataset

04 Pipeline

05 Results

01 CARBON FOOTPRINT

- Cloud computing is the largest GHG emitter (~2.5-3.7% of global)
- “Green AI”: quantify & minimize carbon footprint of AI/ML models
- Build climate resilient AI of use in an energy constrained future

“Quantifying the Carbon Emissions of Machine Learning” (2019)

→ Machine Learning Emissions Calculator + Guidelines

“Towards Climate Awareness in NLP Research” (2022)

→ Climate Performance Model Card

“Estimating the Carbon Footprint of BLOOM” (2022)

→ Training = 50 tons of CO_2 = 60 flights from London to NYC

01 CARBON FOOTPRINT

Factors of carbon footprint:

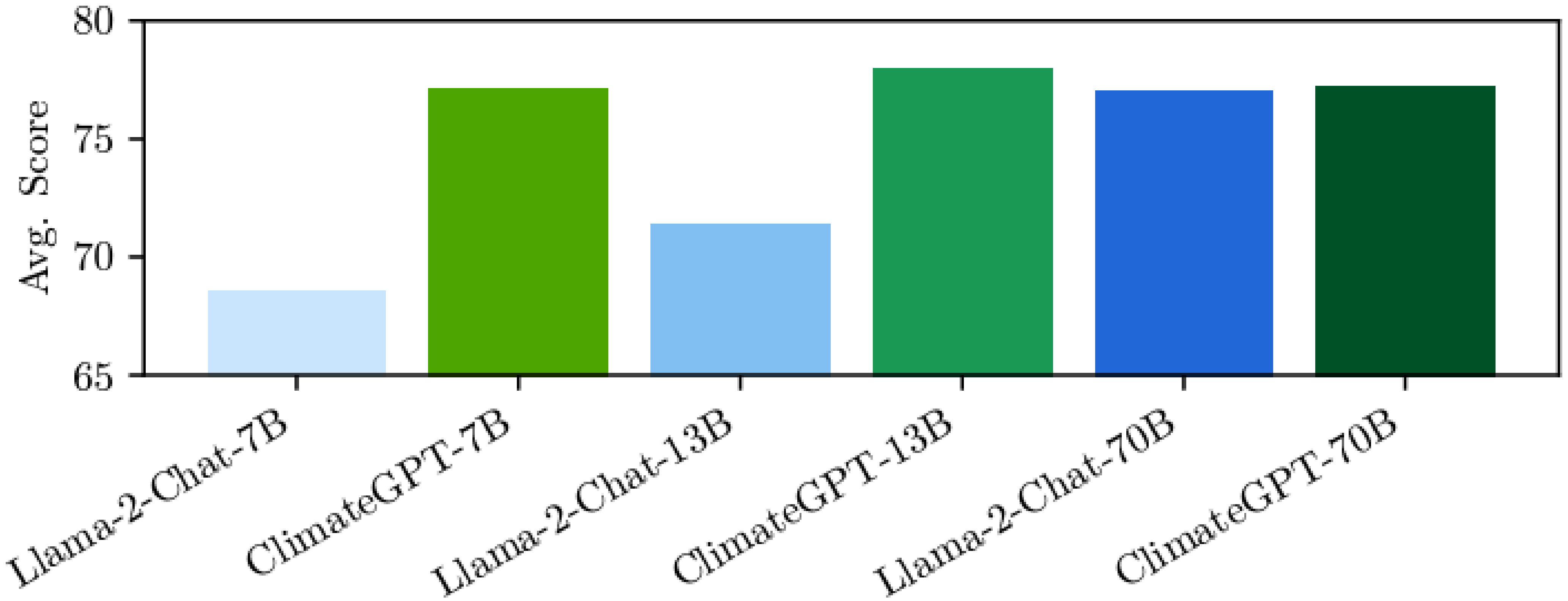
- Server: carbon intensity + time of day
- Runtime: model size + execution frequency
- Hardware: efficiency + resource allocation

Be informed & intentional

- Transparent tracking and reporting of emissions
- Consciously choose compute infrastructure and location
- Minimize unnecessary training (OTS models, tuning)
- Random and selective hyperparameter search >> gridsearch

02 CLIMATE GPT

- The Endowment for Climate Intelligence, "ClimateGPT: Towards AI Synthesizing Interdisciplinary Research on Climate Change"
- Llama 2 backbone + continued pretraining on 4.2B curated climate tokens + instruction finetuning
- Climate specific model intended for OTB use and finetuning



03 DATASET

ClimateBERT Climate Sentiment

- 1000 train/320 test [text, label] pairs
- Sentiment labels: {0: “risk”, 1: “neutral”, 2: “opportunity”}

Preprocessing: 800 train/200 val/320 test ChatML prompts

```
{"messages": [
    {"role": "system", "content": "You are..."},
    {"role": "user", "content": "..."},
    {"role": "assistant", "content": "..."}
]}
```

```python
def create_prompt(sample):
  return {
    "messages": [
      {"role": "system", "content": SYSTEM_PROMPT},
      {"role": "user", "content": sample["text"]},
      {"role": "assistant", "content": IDX2LBL[sample["label"]]}
    ]
  }
```

```python
SYSTEM_PROMPT = """
  Analyze the sentiment of the user provided content and determine if
  it is describing risk, opportunity, or neutral sentiment related to
  climate and the environment. Your response should be the corresponding
  sentiment label "risk" or "opportunity" or "neutral".
"""
```

04 PIPELINE

Compute Infrastructure: AWS & Electricity Maps

- Instance: AWS g5.16xlarge, Ubuntu 20.04, 64 vCPUs, 256 GiB memory
- Single GPU: NVIDIA A10G Ampere Tensor Core (24 GiB)
- Location: Virginia ("us-east-1") = 100% renewable energy (solar)

Finetuning Pipeline: Huggingface

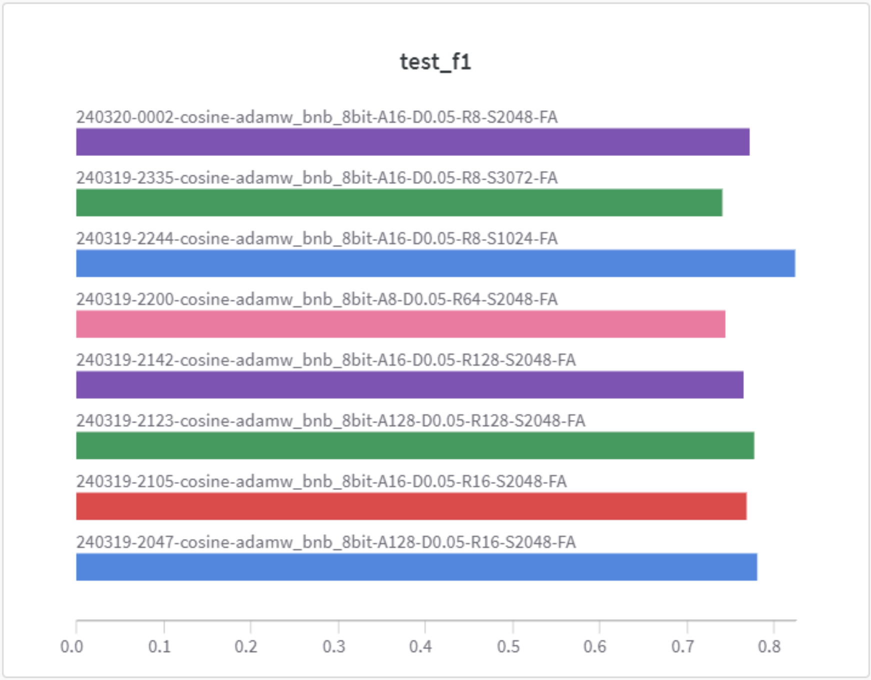
- PEFT: BitsandBytes + QLoRA
- TRL: Supervised Finetuning with ChatML prompts

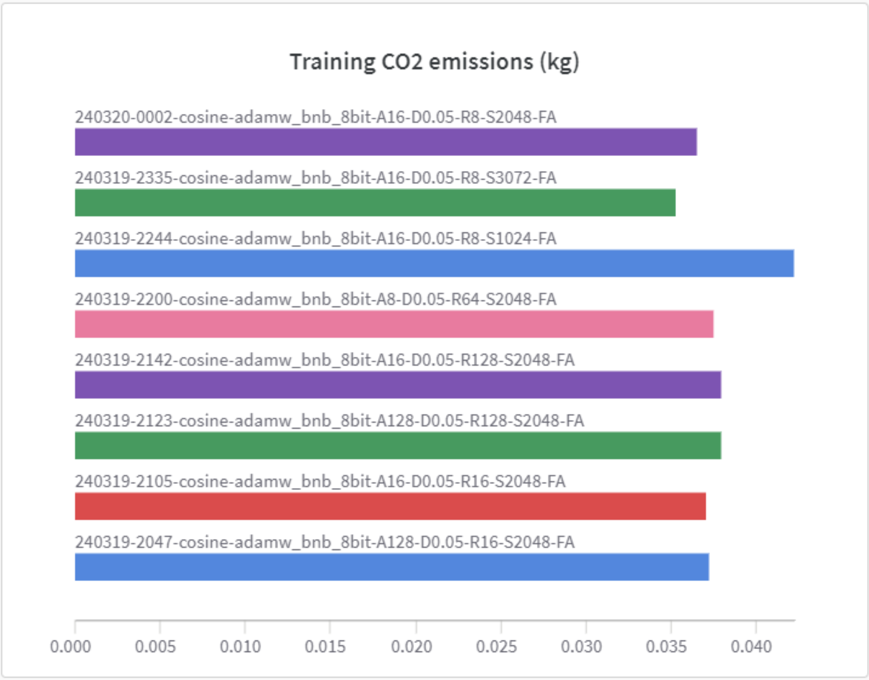
04 PIPELINE

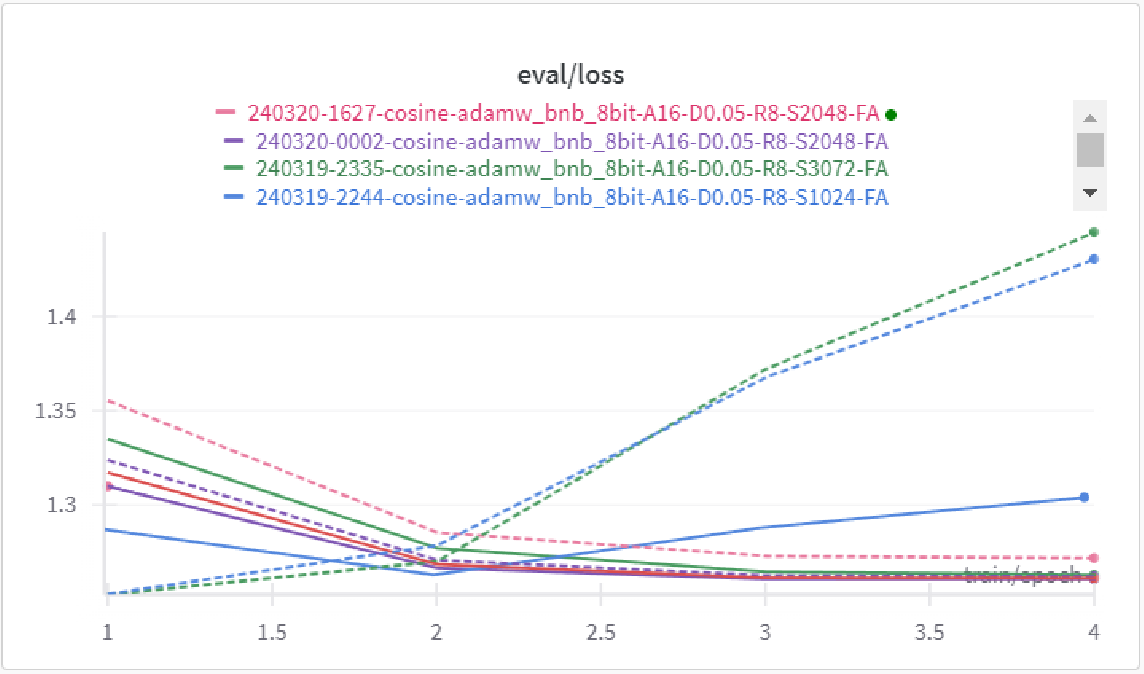
Emissions Tracking: CodeCarbon

Model Tuning & Logging: Weights&Biases

- Random/selective hyperparameter sweep over alpha, rank, max seq
- “Practical Tips for Finetuning LLMs Using LoRA (Low-Rank Adaptation)”
- HF docs: “Methods and tools for efficient training on a single GPU”



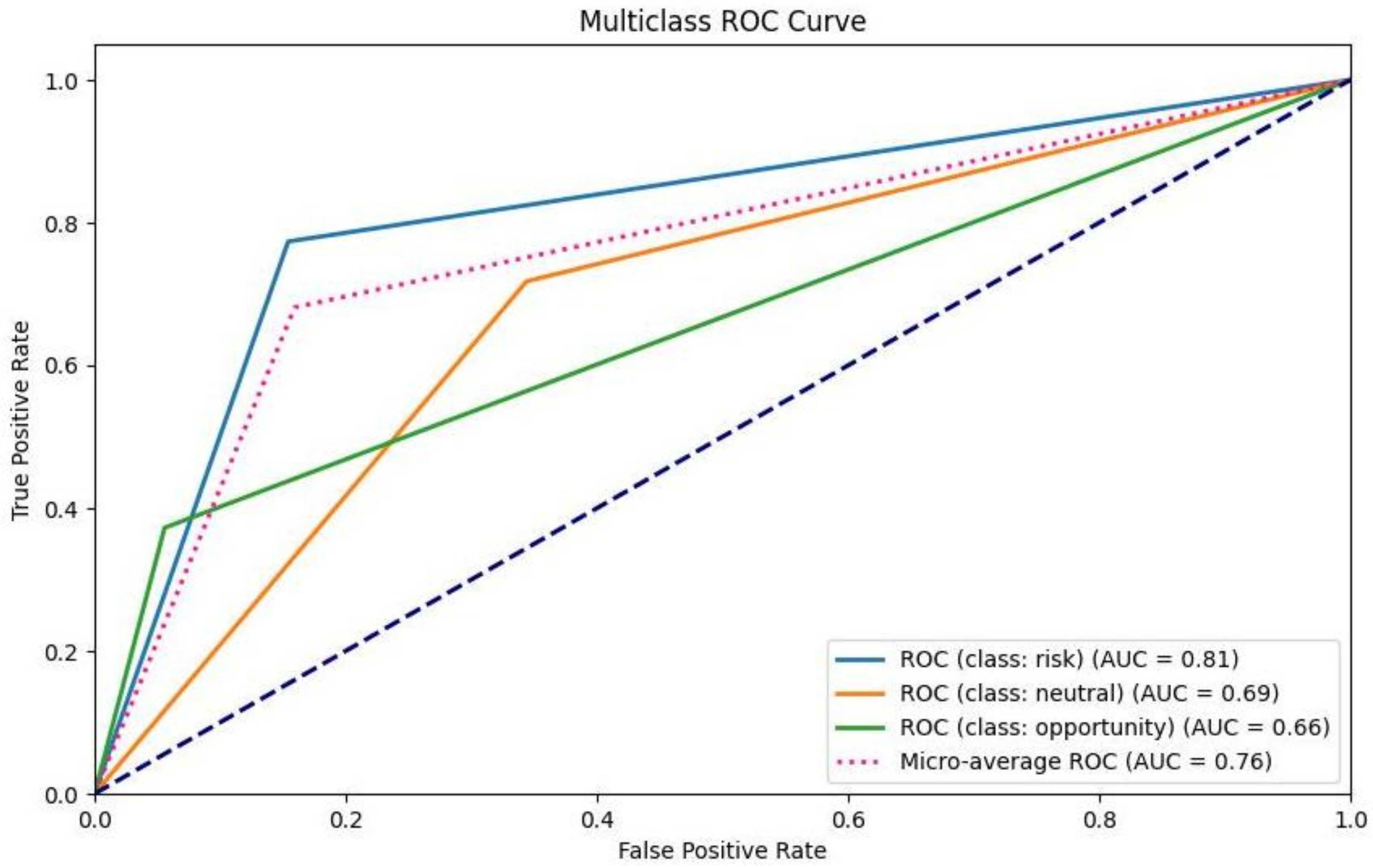




05 RESULTS

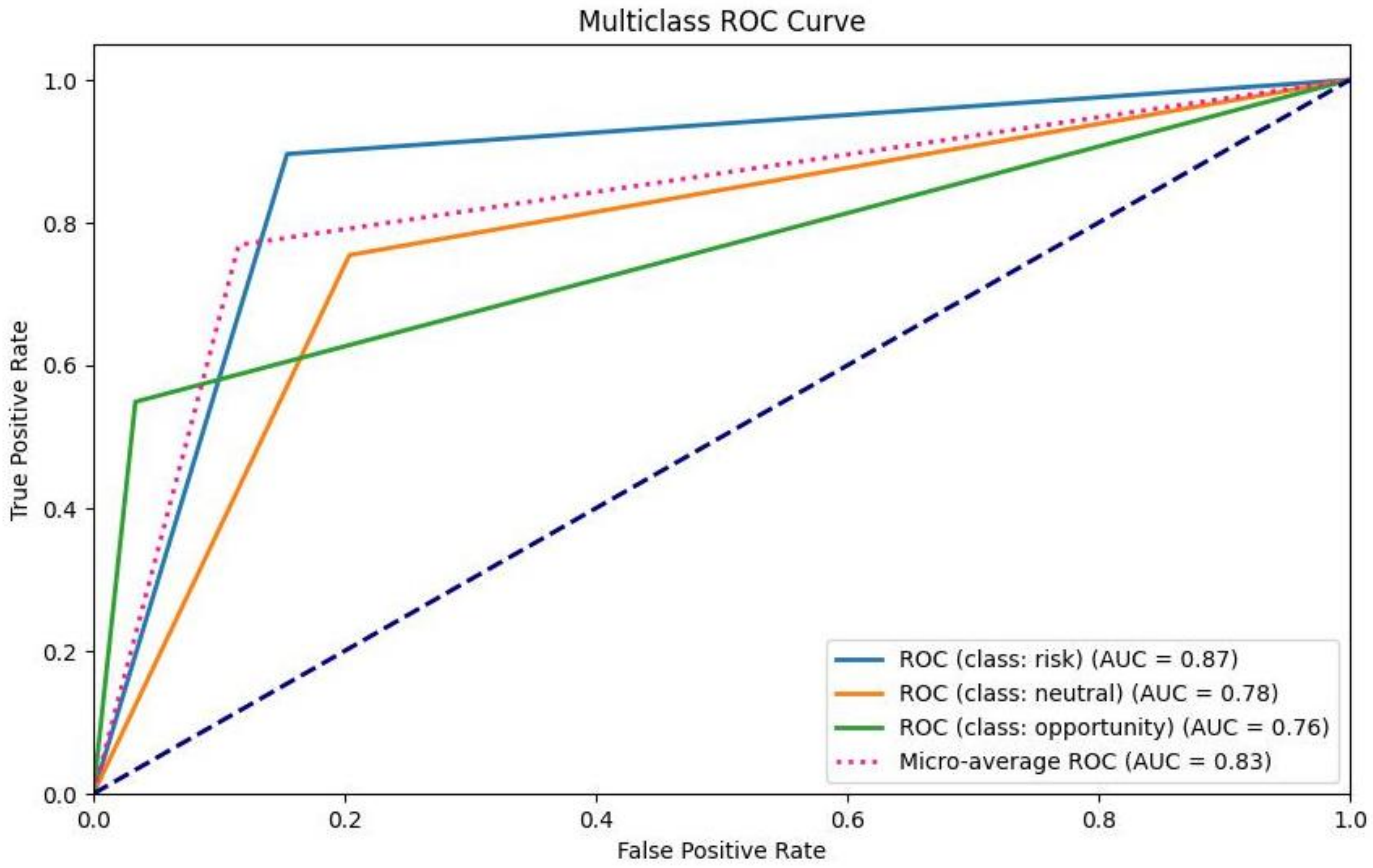
eci-io/climategpt

```
Test recall: 0.62
Test precision: 0.65
Test f1: 0.63
Test accuracy: 0.68
```



kahliahogg/climate-peft

```
Test recall: 0.73
Test precision: 0.76
Test f1: 0.74
Test accuracy: 0.77
```



05 RESULTS

Climate Performance Model Card	eci-io/climategpt	kahliahogg/climate-peft
1. Is the resulting model publicly available?	Yes	Yes
2. Time to train final models	64,500 GPU hours	0.28 GPU hours
3. Time for all experiments	3685 GPU hours	2.27 (16.5) GPU hours
4. Energy consumption GPU + CPU + RAM	0.78 kW	0.102 kW
5. Geolocation for computation	Washington, USA	Virginia, USA
6. Energy mix at the geolocation	24 gCO2eq/kWh	335 gCO_2eq/kWh
7. CO_2eq emissions to train the final model	1,199.70 kg	0.04 kg
8. CO_2eq emissions for all experiments	333 kg	0.30 kg
9. Average CO_2eq emission per inference sample	6.6e-05 kg	6.0e-06 kg