

Configuration Manual

MSc Research Practicum
MSc in Artificial Intelligence

Nagalakshmi Guntumadugu
Student ID: 23280409

School of Computing
National College of Ireland

Supervisor: Ade Fajemisin

National College of Ireland
MSc Project Submission Sheet



School of Computing

Student Name: Nagalakshmi Guntumadugu

Student ID: 23280409

Programme: MSc in Artificial Intelligence

Year: 2024-2025

Module: MSc Research Practicum

Lecturer: Ade Fajemisin

Submission

Due Date: 15/09/2025

Project Title: EmotiAug: Enhancing Minority Emotion Classification Through Chain-of-Thought Data Augmentation

Word Count: 367 **Page Count:** 5

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Nagalakshmi Guntumadugu

Date: 14/09/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	☐
Attach a Moodle submission receipt of the online project submission , to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Nagalakshmi Guntumadugu
23280409

Emotion Augmentation & Training Configuration Manual

What is This?

DeepSeek-R1 Emotion Augmentation System - Generates synthetic emotional text samples for minority emotion classes to improve BERT model performance on imbalanced datasets.

Problem: Some emotions (grief, pride, fear) have very few training samples

Solution: Generate high-quality samples using DeepSeek-R1 → Train better models

1 4 Notebooks Overview

1.1 Data Augmentation (2 Notebooks)

1. **NagaLakshmiFirstDataset.ipynb** → Generate samples for first dataset
2. **NagaLakshmiSecondDataset.ipynb** → Generate samples for second dataset

1.2 Training & Evaluation (2 Notebooks)

3. **TrainingGoEmotions.ipynb** → Train BERT on GoEmotions dataset
4. **TrainingDAIR.ipynb** → Train BERT on DAIR-AI dataset

2 Augmentation Notebooks - How to Run

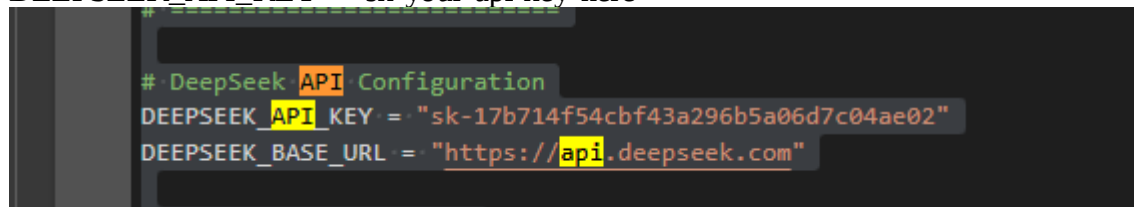
2.1 Step 1: Upload

Upload NagaLakshmiFirstDataset.ipynb to Google Colab

2.2 Step 2: Change API Key

Find this line and change it:

DEEPSEEK_API_KEY = "sk-your-api-key-here"

A screenshot of a code editor with a dark background. It shows a code block for DeepSeek API configuration. The first line is a comment: "# Find this line and change it:". The second line is "DEEPSEEK_API_KEY = \"sk-your-api-key-here\"". The third line is a comment: "# DeepSeek API Configuration". The fourth line is "DEEPSEEK_API_KEY = \"sk-17b714f54cbf43a296b5a06d7c04ae02\"". The fifth line is "DEEPSEEK_BASE_URL = \"https://api.deepseek.com\"". The text "API" in the comments and the API key and URL are highlighted in yellow in the original image.

```
# Find this line and change it:
DEEPSEEK_API_KEY = "sk-your-api-key-here"

# DeepSeek API Configuration
DEEPSEEK_API_KEY = "sk-17b714f54cbf43a296b5a06d7c04ae02"
DEEPSEEK_BASE_URL = "https://api.deepseek.com"
```

2.3 Step 3: Change Paths

Find these lines and update paths:

DATASET_PATH = "/content/your_dataset.csv"

SAVE_DIR = "/content/drive/MyDrive/YourFolder/"

```
# Dataset Configuration
DATASET_PATH = "/content/goemotions_clean.csv"
```

```
# Mount drive
from google.colab import drive
drive.mount('/content/drive')
```

3 Step 4: Run

```
# For testing (recommended first):
results = test_with_response_inspection(['fear', 'desire'])
```

```
# For production:
results = run_full_augmentation_by_priority(1) # Priority 1 emotions
```

```
# OR
results = run_full_augmentation_by_priority(2) # Priority 2 emotions
```

```
# Test all P2 emotions (fear, desire, disgust, surprise, sadness)
results = test_all_p2_emotions()
```



Testing all Priority 2 emotions...

```
# OR
results = run_full_augmentation_by_priority() # All emotions
Repeat same steps for NagaLakshmiSecondDataset.ipynb
```

4 Training Notebooks - How to Run

4.1 Step 1: Upload

Upload TrainingGoEmotions.ipynb to Google Colab

4.2 Step 2: Change Dataset Paths

Find these lines and update:

```
ORIGINAL_DATASET = "/content/goemotions_clean.csv"
AUGMENTED_DATA = "/content/drive/MyDrive/YourFolder/combined_final_xxxxx.csv"
MODEL_SAVE_DIR = "/content/drive/MyDrive/Models/"
```

4.3 Step 3: Run

Just run all cells - the notebook handles:

```
# - Loading data
# - Training baseline model (original data)
# - Training augmented model (original + augmented)
# - Comparing performance
```

Repeat same steps for TrainingDAIR.ipynb

5 Quick Checklist

5.1 Augmentation Phase

- Upload augmentation notebook
- Set DEEPSEEK_API_KEY = "sk-your-key"
- Set DATASET_PATH = "/path/to/dataset.csv"
- Set SAVE_DIR = "/path/to/save/"
- Run: test_with_response_inspection(['fear', 'desire'])
- Run: run_full_augmentation_by_priority(1)

6 Training Phase

- Upload training notebook
- Set ORIGINAL_DATASET = "/path/to/original.csv"
- Set AUGMENTED_DATA = "/path/to/augmented.csv"
- Set MODEL_SAVE_DIR = "/path/to/models/"
- Run all cells
- Check F1 improvement results

7 Runtime

- **Augmentation:** 2-4 hours per notebook
- **Training:** 1 hour per notebook

8 Expected Files

After Augmentation:

/YourFolder/combined_final_XXXXXX.csv

After Training:

/Models/baseline_model/

/Models/augmented_model/

That's it! Upload → Change paths → Run

Outputs

```
# Test all P2 emotions (fear, desire, disgust, surprise, sadness)
results = test_all_p2_emotions()

Parallel processing: 1x
TEST MODE: Limited to 50 samples

Round 1:
Generated: 0
Target: 35 (70%)
Minimum: 30 (60%)
Round 1: 0% | 0/50 [00:00<?, ?it/s]2025-07-07 03:51:29,499 - EmotiAug - INFO - Attempt 1: Generating 15 samples in 1 batches
2025-07-07 03:51:29,499 - EmotiAug - INFO - Attempt 1: Generating 15 samples in 1 batches
2025-07-07 03:51:29,499 - EmotiAug - INFO - Attempt 1: Generating 15 samples in 1 batches
INFO:EmotiAug:Attempt 1: Generating 15 samples in 1 batches
DEBUG:EmotiAug:Filtering 15 samples for fear/deep_emotional_expression (min_score: 13)
2025-07-07 03:53:05,097 - EmotiAug - INFO - Rejection reasons for fear/deep_emotional_expression:
2025-07-07 03:53:05,097 - EmotiAug - INFO - Rejection reasons for fear/deep_emotional_expression:
2025-07-07 03:53:05,097 - EmotiAug - INFO - Rejection reasons for fear/deep_emotional_expression:
INFO:EmotiAug:Rejection reasons for fear/deep_emotional_expression:
2025-07-07 03:53:05,100 - EmotiAug - INFO - lacks_genuine_emotion: 4
2025-07-07 03:53:05,100 - EmotiAug - INFO - lacks_genuine_emotion: 4
2025-07-07 03:53:05,100 - EmotiAug - INFO - lacks_genuine_emotion: 4
2025-07-07 03:53:05,100 - EmotiAug - INFO - lacks_genuine_emotion: 4
INFO:EmotiAug: lacks_genuine_emotion: 4
2025-07-07 03:53:05,103 - EmotiAug - INFO - physical_emotional: 2
2025-07-07 03:53:05,103 - EmotiAug - INFO - physical_emotional: 2
2025-07-07 03:53:05,103 - EmotiAug - INFO - physical_emotional: 2
2025-07-07 03:53:05,103 - EmotiAug - INFO - physical_emotional: 2
INFO:EmotiAug: physical_emotional: 2
2025-07-07 03:53:05,106 - EmotiAug - INFO - word_count: 1
2025-07-07 03:53:05,106 - EmotiAug - INFO - word_count: 1
2025-07-07 03:53:05,106 - EmotiAug - INFO - word_count: 1
2025-07-07 03:53:05,106 - EmotiAug - INFO - word_count: 1
INFO:EmotiAug: word_count: 1
2025-07-07 03:53:05,109 - EmotiAug - INFO - Quality filter: 8/15 passed (53.3%)
2025-07-07 03:53:05,109 - EmotiAug - INFO - Quality filter: 8/15 passed (53.3%)
2025-07-07 03:53:05,109 - EmotiAug - INFO - Quality filter: 8/15 passed (53.3%)
```

```
Q Commands + Code + Text ▶ Run all ▼

INFO:EmotiAug:Quality threshold: 12 (adaptive)
2025-07-09 03:58:01,187 - EmotiAug - INFO - Multi-GPU: DISABLED
INFO:EmotiAug:Multi-GPU: DISABLED
Loaded 211,225 samples
Found 28 emotion classes

Selected 1 emotions
Features enabled:
- Chain-of-Thought Reasoning
- Adaptive Quality Scoring
- Multi-GPU Processing: NO
- Parallel Requests: 1x
- Enhanced Extraction Logic

Starting FINAL Enhanced Augmentation Pipeline
Emotions to process: 1
Chain-of-Thought: ENABLED
Parallel Processing: 1x
Adaptive Quality: ENABLED
Multi-GPU: DISABLED

=====
Processing 1/1: EMBARRASSMENT
=====

Augmenting EMBARRASSMENT
Current samples: 1,720
Target augmented: 2,280
Quality samples: 250
Quality threshold: 12 (adaptive)
Parallel processing: 1x

Round 1:
Generated: 0
Target: 1938 (85%)
Minimum: 1596 (70%)
Round 1: 0% | 0/1938 [00:00<?, ?it/s]2025-07-09 03:58:01,577 - EmotiAug - INFO - Attempt 1: Generating 30 samples in 1 batches
INFO:EmotiAug:Attempt 1: Generating 30 samples in 1 batches
DEBUG:EmotiAug:Filtering 30 samples for embarrassment/deep_emotional_expression (min_score: 14)
2025-07-09 03:59:55,564 - EmotiAug - INFO - Rejection reasons for embarrassment/deep_emotional_expression:
INFO:EmotiAug:Rejection reasons for embarrassment/deep_emotional_expression:
2025-07-09 03:59:55,570 - EmotiAug - INFO - lacks_genuine_emotion: 16
INFO:EmotiAug: lacks_genuine_emotion: 16
2025-07-09 03:59:55,571 - EmotiAug - INFO - Quality filter: 14/30 passed (46.7%)
INFO:EmotiAug:Quality filter: 14/30 passed (46.7%)
```

```
love_result = run_production_emotion('love')
```



PRODUCTION RUN: LOVE

Target: 15,000 samples
Estimated time: 75-100 minutes

Using pre-loaded DAIR-AI emotion dataset...

Ready with 416,809 samples
Found 6 emotion classes

STARTING COMPLETE LOVE AUGMENTATION

Target: 15,000 samples
Max parallel power: 150 operations
Quality threshold: 8

ULTRA-FAST LOVE AUGMENTATION

Target: 15,000 samples
Strategies: 10 parallel
Batch size: 100
Parallel requests per strategy: 15
Total concurrent operations: 150
Quality threshold: 8
Tracking: Raw generated → Quality filtered → Final saved
Checkpoints: Every 500 samples + every 4 chunks
Metrics: Success rates, strategy performance, ETA
Quality seed samples: 300
Samples per strategy: 1500
Batches per strategy: 16
Total batches: 160
Created 160 parallel tasks
Processing in 11 chunks of 15

Success rate: 98.8%



Quality Statistics:

Average quality score: 14.4
Average word count: 27.9
High quality samples (12+): 8359
Chain-of-thought samples: 0

Strategy Performance:

narrative_emotional_journey: 1,080 samples
relationship_emotional_dynamics: 1,055 samples
memory_emotional_flashback: 1,052 samples
physical_emotional_manifestation: 1,021 samples
sensory_emotional_experience: 997 samples
deep_emotional_expression: 977 samples
conversational_emotional_revelation: 947 samples
internal_emotional_monologue: 921 samples
metaphorical_emotional_landscape: 911 samples
environmental_emotional_context: 869 samples

Results saved:

Samples: /content/drive/MyDrive/DAIR_AI_UltraFast_Augmentation/love_9830_20250720_165833.csv
Statistics: /content/drive/MyDrive/DAIR_AI_UltraFast_Augmentation/love_stats_20250720_165833.json

Sample Examples:

1. (Score: 19) My heart flutters wildly when I see your sleepy smile in the morning sunlight. I deeply cherish thes...
2. (Score: 19) Just the thought of your laugh sends warm tingles down my spine. I adore how you scrunch your nose; ...
3. (Score: 19) Your gentle kiss on my forehead last night made my whole body tingle. It was such a simple act of ca...

PRODUCTION COMPLETE!

Success: 9,830 / 15,000 samples
Time: 81.6 minutes
Saved to: /content/drive/MyDrive/DAIR_AI_UltraFast_Augmentation/love_9830_20250720_165833.csv