Challenge 3 Report

Vibe Coding: Play to Impact

Team Name: UDU_FarmCookie

Institution: Posts and Telecommunications

Institute of Technology

Date: 20 October 2025

Table of Contents:

1. Introduction

Background and Motivation

Cyberbullying has become a critical issue in the digital age, affecting approximately 59% of teenagers worldwide. Traditional reactive approaches like reporting systems are insufficient to address this growing problem. There is a need for proactive educational tools that help users recognize harmful language and develop empathetic communication skills.

Problem Statement

Current anti-cyberbullying initiatives lack interactive, engaging methods to teach young people about identifying and transforming negative online behavior. Educational content is often static and fails to provide personalized learning experiences.

Project Objectives

Empathy Defender is an AI-powered educational game designed to:

- Teach users to distinguish between negative and positive online communication
- Develop critical thinking skills in analyzing emotional content
- Practice transformation of harmful language into empathetic responses
- Raise awareness about the impact of words on mental health
- Leverage AI for dynamic, personalized learning experiences

Target Audience

Primary audience includes teenagers and young adults (ages 13-25), with secondary reach to educators, parents, and social workers interested in digital literacy and online safety.

2. Game Overview

Game Concept

Players assume the role of an "Empathy Defender" who fights cyberbullying through kindness. The game presents AI-generated social media comments that players must identify as positive or negative, then transform negative comments into positive ones.

Core Gameplay Loop

Setup: Players enter their OpenAI API key

Story Mode: Optional narrative introduction setting mission context

Classification: Identify comments as positive or negative (10 points)

Transformation: Rewrite negative comments positively (20 points)

Feedback: AI analyzes responses and provides immediate feedback

Progression: Complete 10 rounds with increasing difficulty

Key Features

AI-Powered Content: Dynamic generation of realistic social media comments

Intelligent Analysis: Deep semantic understanding beyond keyword matching

Gamification: Scoring system with streak multipliers (2x, 3x, etc.)

Visual Feedback: Animated emojis and color-coded responses

Educational Focus: Tips, guidance, and immediate learning opportunities

User Experience

The game uses a warm color palette (pink/purple gradients) to evoke empathy, with a dark theme for comfortable extended play. Navigation flows from setup through menu, optional story mode, gameplay, and results screen. The interface is responsive and accessible across all devices.

3. Game Mechanics and Design

Game States

Setup State: Secure API configuration with privacy assurance

Menu State: Game orientation showing objectives and rules

Story State: Five-part narrative with typewriter animation

Playing State: Core gameplay with classification and transformation

Result State: Performance summary with replay option

Scoring Mechanism

Correct Classification: 10 base points

Successful Transformation: 20 base points

Streak Bonus: Base points × streak multiplier

Example: 3-streak classification = 10 + (10 × 3) = 40 points

Difficulty Progression

AI generates varied content that naturally increases complexity:

Early rounds: Clear positive/negative distinctions

Middle rounds: Subtle sarcasm and context-dependent meanings

Later rounds: Complex statements requiring deeper analysis

Feedback Systems

Visual: Color-coded borders (green for correct, red for incorrect), animated floating emojis, loading spinners

Textual: Success messages with point totals, constructive error guidance, contextual tips

Accessibility: High contrast, keyboard navigation support, clear icons with text labels

4. Technical Implementation

Technology Stack

Frontend: React 18+ with TypeScript, Next.js

Styling: Tailwind CSS for responsive design

State Management: React Hooks (useState, useRef, useEffect)

External API: OpenAI GPT-3.5-turbo

Icons: Lucide React library

Component Architecture

Main component EmpathyDefender manages game state, score, streak, rounds, user input, loading states, and feedback messages. Key state variables control screen display, API credentials, gameplay mechanics, and UI interactions.

Core Functions

generateMessageWithAI(): Creates realistic social media comments using GPT-3.5

analyzeEmotionWithAI(): Classifies text as positive or negative using AI

handleClassification(): Processes player's classification choice with locking mechanism

transformToPositive(): Validates player's transformation attempt

addEmotionAnimation(): Triggers floating emoji particles

Security and Privacy

API keys stored client-side only (session memory)

No data sent to external servers except OpenAI

Comprehensive error handling with clear messages

Input validation and duplicate request prevention

Performance Optimization

Conditional rendering based on game state

Button locking prevents rapid-fire submissions

Minimal re-renders with batched state updates

Tree-shakeable icon library and inline CSS

5. AI Integration and Development Process

AI Technology Selection

OpenAI GPT-3.5-turbo chosen for:

Semantic understanding of context, tone, and implied meanings

Natural language generation for authentic comments

Flexible prompt-based customization

Simple REST API with free tier availability

Reliable classification with proper prompt engineering

Prompt Engineering

Classification System Prompt: "You are an emotion analysis expert. Analyze whether a statement is negative, violent, offensive, or bullying. Reply only with 'NEGATIVE' or 'POSITIVE'."

Generation Prompts: Create concise 10-15 word social media comments (educational context), with temperature 0.8 for variety and 0.3 for classification consistency.

Development Iterations

Rule-Based Prototype: Used keyword lists, failed with contextual negativity

Sentiment Library: Struggled with sarcasm and cultural context

OpenAI Integration: Successfully handles nuance, generates infinite content

Testing and Validation

Comprehensive testing for sarcasm, backhanded compliments, cultural phrases, and ambiguous statements. Fallback mechanism displays transparent error messages without silent failures.

Ethical Considerations

Age-appropriate content generation

No extreme or harmful language

Bias monitoring during development

Complete privacy protection (no data collection)

Cost Analysis

Per game session (10 rounds): ~2,500 tokens = $0.004. Scalability achieved through player-provided API keys, distributing costs without central billing.

6. Impact and Relevance

Educational Impact

Learning Outcomes: Students understand cyberbullying forms, recognize subtle negativity, analyze emotional content, reframe negative statements, and develop empathy.

Pedagogical Approach: Constructivist learning through active interaction, game-based motivation with safe practice environment, spaced repetition across 10 rounds.

Social Relevance

Addresses critical issue where 59% of teens experience online harassment. Traditional education lacks engaging methods; this game provides interactive, personalized learning that reaches digital natives in their preferred format.

Practical Applications

Schools: Integrate into digital citizenship curriculum

Counseling: Tool for discussing online behavior and emotional intelligence

Workshops: Interactive component for anti-bullying programs

Self-Learning: Accessible resource for independent skill development

Measurable Benefits

Players improve recognition accuracy, develop positive communication patterns, increase empathy awareness, and gain confidence in addressing online negativity. The scoring system provides quantifiable progress tracking.

7. Reflection and Future Work

Project Achievements

Successfully created an engaging educational game that leverages AI for dynamic content generation and analysis. Achieved seamless integration of GPT-3.5-turbo with intuitive user interface. Implemented comprehensive gamification that maintains player motivation across multiple sessions.

Lessons Learned

Technical: Prompt engineering critical for AI reliability; client-side API key approach balances cost and privacy. Design: Simple mechanics more effective than complexity; immediate feedback essential for learning. Educational: Narrative context increases engagement; mistakes as learning opportunities reduces anxiety.

Limitations

Requires OpenAI API key (barrier for some users)

AI classification occasionally inconsistent with human judgment

Limited to text-based interactions (no image/video analysis)

English language only

No progress persistence across sessions

Future Enhancements

Technical Features: Multiplayer mode for collaborative learning, achievement system with badges, difficulty levels (beginner/advanced), persistent user profiles with progress tracking

Content Expansion: Multi-language support, custom comment categories (racism, sexism, etc.), real-world case studies, video/image content analysis

Educational Tools: Teacher dashboard for classroom use, detailed analytics and reports, customizable learning paths, parent monitoring features

Community Features: Share transformations with friends, leaderboard system, community-created content, discussion forums for best practices

Scalability Roadmap

Short-term: Implement local caching, add more visual themes. Mid-term: Develop backend for shared API keys in educational settings, create mobile app version. Long-term: Partner with schools for pilot programs, integrate with LMS platforms, expand to corporate training for workplace communication.

8. References

Academic Sources

Pew Research Center (2022). "Teens, Social Media and Technology 2022"

CDC Youth Risk Behavior Survey (2021). "Cyberbullying and Mental Health"

Kowalski, R.M. et al. (2014). "Bullying in the digital age: A critical review"

Hinduja, S. & Patchin, J.W. (2020). "Cyberbullying Identification, Prevention, and Response"

Technical Documentation

OpenAI API Documentation. https://platform.openai.com/docs

React Documentation. https://react.dev

Next.js Documentation. https://nextjs.org/docs

Tailwind CSS Documentation. https://tailwindcss.com/docs

Design and UX

Deterding, S. et al. (2011). "Gamification: Toward a Definition"

Kapp, K.M. (2012). "The Gamification of Learning and Instruction"

Nielsen Norman Group. "Usability Heuristics for User Interface Design"

AI and NLP

Brown, T. et al. (2020). "Language Models are Few-Shot Learners" (GPT-3 paper)

Liu, P. et al. (2023). "Pre-train, Prompt, and Predict: A Systematic Survey"

Bommasani, R. et al. (2021). "On the Opportunities and Risks of Foundation Models"

Online Safety and Digital Citizenship

Common Sense Media. "Digital Citizenship Curriculum"

UNICEF (2019). "Growing up in a connected world"

Cyberbullying Research Center. https://cyberbullying.org