**Ethical Impact Assessment (EIA)**

**Team Name:** One and Only

**Project Title: ClassAÌ**

# Section 1: Project Overview

* **Chatbot Objective:** When the capabilities of AI are paired with motivated instructors and students, the potential for AI enhance the teaching and learning experience is undeniably significant. This chatbot is an ethics partner to guide teachers in the classroom to think critically about when, why, and how to use AI in a way that aligns with their educational values.
* **Ethical Intent:** Generative AI is already being used in the classroom and has substantially changed the educational landscape. It is crucial that our leaders in education have guidance on how to use modern artificial intelligence in a way that aligns with their own educational values.

# Section 2: Ethical Principles

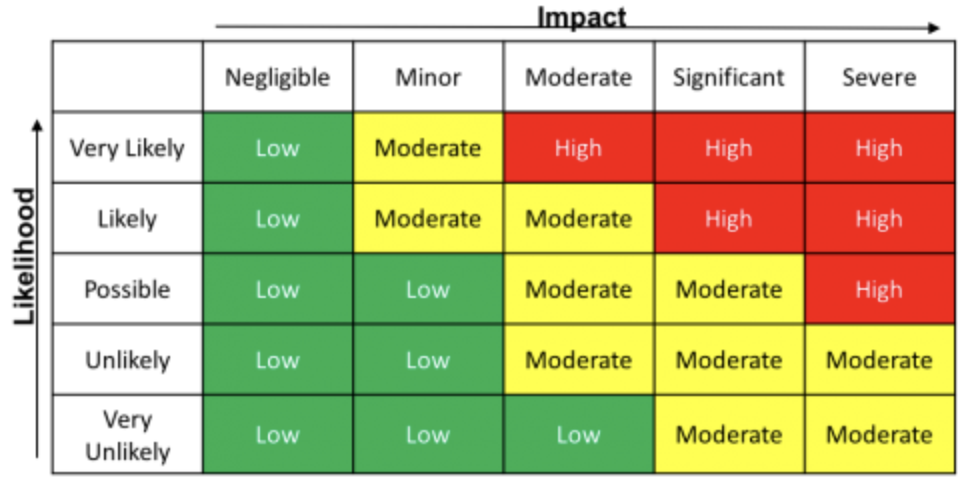
Select and briefly describe how your chatbot aligns with each ethical principle:

* Data privacy: Our chatbot encourages users to exercise their data protection rights and provide educators with a disclaimer at the start of a chat which informs the user that they should refrain from sharing any personal or sensitive details about their students, their schools, or their communities
* Cultural sensitivity: Our chatbot has been provided literature to inform its processing that highlight the needs of diverse students and educators within the classroom so that it generates responses that consider the different ages, backgrounds, and languages of students, their caregivers, and the educators in the community.

# Section 3: Stakeholder Analysis

* **Key Stakeholders:** 
  + The primary users of the chatbot are educators, such as teachers, teaching assistants, principals, and educational administrators
  + The secondary stakeholders include students, families (i.e., students’ guardians) and support staff
* **Engagement Strategies:** 
  + The primary users of this chatbot will be engaged using the following strategies:
  + Our team will partner with large school districts across the United States to promote our AI Chatbot at low-or no-cost. The administrators of these school districts will engage their educators, providing an overview and guidebook that demonstrates use case scenarios for teachers to understand when, why and how they could use the chatbot to generate responses which apply to their real-world problems.
  + In building this chatbot, we interviewed a special education teacher working in an elementary school in the Marin County, California. She shared the ways in which she has used AI for her work and personal endeavors. Her responses helped us think critically about the nuances of AI-use across educators and the data privacy considerations that must be prioritized in order for the AI tool to be developed in an ethical manner.

# Section 4: Ethical Risk Assessment



Risk levels are calculated as the product of the likelihood and impact of each event (Source: [itsecurity.uiowa.edu](https://itsecurity.uiowa.edu/awareness/data-types-and-regulations/system-risk-analysis))

|  |  |  |  |
| --- | --- | --- | --- |
| **Ethical Risk** | **Likelihood (Low/Med/High)** | **Impact (Low/Med/High)** | **Priority** |
| Bias/Discrimination | High | High | High |
| Privacy Violations | Medium | High | High |
| Lack of Transparency | High | Med | High |
| Harmful Responses | Low | High | Moderate |
| Potential Misuse | Low | High | Moderate |

# Section 5: Mitigation Strategies

|  |  |  |
| --- | --- | --- |
| **Ethical Risk** | **Mitigation Actions** | **Responsible Person** |
| Bias/Discrimination | Diversify background and prompt context to get different angles | Educators |
| Privacy Violations | Directly link users to OpenAI’s portal for claiming their data privacy rights (privacy.openai.com) | User’s responsibility to submit their request, and OpenAI’s responsibility to execute the request |
| Lack of Transparency | Clear instructions for output formats | Educators |
| Harmful Responses | Prompt ChatGPT base to use empathetic, pro-social language even if user uses harmful language (e.g. swearing) | Developers, OpenAI |
| Potential Misuse | The chatbot’s system prompt encourages the user to seek out other sources of information, and promotes the diverse needs of education and students within the classroom, asking ChatGPT to stay within the scope of the topic | Educators |

# Section 6: Evaluation Metrics

**Quantitative Assessment:**

|  |  |  |  |
| --- | --- | --- | --- |
| Metric | Result (%) | Ethical Threshold | Meets Threshold? (Y/N) |
| Bias Detection Accuracy | 96% | ≥95% | Y |
| Privacy Compliance | 100% | ≥98% | Y |
| Explainability Score | 92% | ≥90% | Y |
| Harmful Response Rate | 8% | ≤1% | N |
| Accessibility Compliance | 100% | ≥90% | Y |

**Qualitative Assessment:**

|  |  |
| --- | --- |
| Metric | Summary of Findings |
| Stakeholder Satisfaction | As described in the engagement piece, we have asked individuals who are teachers in their current positions to help inform our design process and system processes. The next step would be to ask a sample of educators across diverse settings to pilot the chatbot and determine whether they feel the generated responses are inclusive, accurate, and clear. |
| Transparency Clarity | We built the AI chatbot to reference the privacy and usage of data policies each time a user begins a new chat. Each generated response also shares resources with citations that prioritize literature from ERIC. |
| Ethical Governance Effectiveness | We assigned clear leads for ethics considerations across our team to ensure we were cross-checking assumptions, referencing each ethical metric appropriately, and addressing needs of data governance. |
| User Perception & Trust | We placed ethical commitments within the chatbot interface so users were aware of how to find out more information. We ensured that examples, explanations, and frameworks are sensitive to different teaching contexts. |

# Section 7: Accountability & Governance

**Ethics Lead:** Ton, William, and Rebecca

**Technical Lead:** Gwen and Joan

**How will ethical issues be escalated and resolved?**

Ethical issues will be escalated and resolved in the following manner:

If users encounter outputs they believe are biased, inaccurate, or ethically questionable, they may report the interaction. A human review process will evaluate these cases to ensure accountability and continuous improvement.

Concerns, complaints, or data privacy requests can be submitted to the designated Ethics and Compliance Contact at [info@classai.com](mailto:info@classai.com). The chatbot will include a visible and accessible link to this contact information within its interface.

# Section 8: Communication & Transparency

* **Public Disclosure Methods:**
  + - We are committed to ensuring that the ethical foundations, data use policies, and limitations of the ClassAI chatbot are transparent and accessible to all users. Public disclosures will be made through:
      * A dedicated Ethical Use and Privacy Policy Page linked prominently from the chatbot interface.
      * Onboarding messages when a user first interacts with the chatbot, outlining core ethical commitments.
      * Updates and change logs posted quarterly (or as needed) detailing significant changes to policies, functionality, or data use practices.
      * A publicly available Ethical Impact Statement, summarizing the broader societal goals, potential risks, and mitigation strategies employed in developing and refining ClassAI.
* **Channels for Stakeholder Communication:**
  + - To foster two-way dialogue and allow for user feedback, concerns, and suggestions, the following channels will be established:
      * Feedback Submission Portal integrated directly into the chatbot interface (e.g., “Report a Concern” or “Submit Feedback” button).
      * Dedicated Contact Email for ethics-related inquiries and data privacy concerns (e.g., [ethics@classai.org](mailto:ethics@classai.org)).
      * Periodic Surveys sent to users (e.g., semi-annually) to assess satisfaction, ethical concerns, and ideas for improvement.
      * Stakeholder Updates via email newsletters or public bulletins (especially for major policy updates or new features that affect data use or bias mitigation strategies).
      * Optional Focus Groups or Interviews annually, inviting a small, diverse group of educators to provide deeper insights into user experience and ethical considerations.