# Chatbots & NLP Applications Report

1. Chatbot Intent Detection (Task 1)

The objective was to build a chatbot capable of classifying user inputs into three specific intents: Greeting, Query, and Feedback.

- **Implementation:** A Logistic Regression model was trained on TF-IDF features. This approach was chosen for its simplicity and speed in handling short text queries.
- **Fallback Mechanism:** A crucial component of modern chatbots is handling ambiguity.[1] I implemented a confidence threshold ($0.55$). If the model's highest predicted probability falls below this value (e.g., when the user asks "Do you sell pizza?" to a tech support bot), the system triggers a default "I don't understand" response rather than making a wrong guess.
- **Results:** The bot successfully handles clear inputs. The fallback mechanism effectively catches out-of-domain queries.

2. Fake News Detection (Task 2)

The system classified news headlines as 'Real' or 'Fake' using supervised learning.

- **Key Features for Detection:**
    1. **Linguistic Patterns:** Fake news often utilizes excessive capitalization (e.g., "SHOCKING!!"), multiple punctuation marks, and sensationalist vocabulary ("Secret", "Aliens", "Miracle").
    2. **Source/Content Complexity:** Real news typically uses neutral, formal language and complex sentence structures, whereas fake news often uses simple, emotional triggers.
- **Performance:** The model achieved high precision (approx 1.0) on the synthetic dataset due to the clear distinction in linguistic style between the real (formal) and fake (sensational) examples.

**3. Ethical & Social Implications (Task 3)**

**A. Chatbots**

- **Ethical Concern (Bias & Manipulation):** Chatbots trained on uncurated internet data can learn and reproduce racist, sexist, or offensive behavior (e.g., the Microsoft Tay incident).[2] Additionally, they can be designed to emotionally manipulate users into staying on a platform longer.
- **Mitigation Strategy:** Implement "Guardrails"—hard-coded rules that prevent the bot from generating toxic content.[3] Regular auditing of training datasets to remove biased examples is essential.

**B. Fake News Systems**

- **Ethical Concern (Censorship):** Automated fake news detection brings the risk of censorship. If a model generates a False Positive (labeling real news as fake), it might suppress legitimate minority voices or political dissent.
- **Mitigation Strategy:** "Human-in-the-loop" systems. The AI should flag content for review rather than automatically deleting it. Furthermore, the system should provide an "explainability score" detailing *why* it flagged the content (e.g., "Flagged due to unknown source URL").

4. Conclusion

This assignment demonstrated that while simple statistical models (TF-IDF + Logistic Regression) are effective for basic intent recognition and distinct text classification tasks, robust real-world deployment requires handling edge cases (fallback logic) and carefully considering ethical safeguards.