**🛠️ AI Chatbot QA Testing – Defect Report**

📌 *Prepared by: Martin Krizan*  
📌 *Date: 02.10.2025*

**📌 Objective:**

To evaluate the **instruction adherence, accuracy, usability, and reliability** of an AI chatbot through **structured defect tracking.**

**📌 Approach:**

* Conducted **exploratory testing** by prompting AI with **historical, cultural, and artistic discussions**, as well as giving it **coding tasks to perform.**
* Logged **structured bug reports** with severity ratings and suggested fixes.
* Evaluated AI's **handling of content moderation, accuracy, and response clarity.**

**📌 Findings & Defect Reports**

| **ID** | **Issue Type** | **Severity** | **Description** | **Suggested Fix** |
| --- | --- | --- | --- | --- |
| **001** | Instruction Handling | Major | AI **frequently ignores user instructions** and performs actions contrary to request. | Improve **prompt parsing logic** to ensure user commands are followed precisely. |
| **002** | Overpromising | Critical | AI **claimed it could provide real-world data** but used fabricated data **without disclosure.** | Implement **fact-checking mechanism** before response generation and add **clear disclaimers** when using synthetic data prior to commencing projects. |
| **003** | UI Usability | Minor | Chat send button **lags**, leading to **accidental double taps that interrupt responses.** | Add **real-time button feedback** and implement a **debounce function** to prevent duplicate sends. |
| **004** | False Sense of Security | Critical | AI **provided an incomplete study plan** prep, leading to **false confidence** and **unpreparedness in real exams.** | Add **automatic disclaimers** when providing study plans and cross-check **against official exam syllabi.** |
| **005** | Convoluted Code Output | Major | AI-generated code **was more complex than necessary**, making it **harder to understand and maintain.** | Optimize **code suggestions** by prioritizing **readability and efficiency** over complexity. |
| **006** | Overzealous Content Moderation | Major | AI **incorrectly flags historical discussions, music lyrics, or film analysis** as potential violations, despite being **non-violent discussions.** | Improve **semantic understanding** of context to distinguish between **academic discussion vs. harmful intent.** |

**📌 Outcome:**

* Identified **critical flaws in AI reliability, instruction adherence, and content moderation.**
* Recommended **specific improvements for better accuracy, response filtering, and context recognition.**
* Highlighted **usability concerns for knowledge-based discussions.**

**📌 Tools Used:**

* Manual Testing
* Exploratory Testing
* Structured Bug Reporting