## 📘 ****1. Project Overview****

**Title**: Adaptive LLM-Based Conversational AI for Varied Personality Types

This project studies how LLMs adapt their dialogue strategies when conversing with different user personalities (oversharers, verbose non-reflective users, reserved users). It emphasizes ethical implications such as over-personalization, manipulation, and reinforcement of bias.

## 🎯 ****2. Objectives****

* Simulate user personalities in controlled dialogue settings.
* Configure or fine-tune LLMs to exhibit adaptive behavior.
* Analyze linguistic, emotional, and strategic adaptation.
* Evaluate ethical and bias risks.
* Develop a working prototype with different user-persona simulations.

## 🧪 ****3. Methodology****

### a. ****Data Collection & Persona Modeling****

* **Resources**:
  + [Big Five Personality Traits (Wikipedia)](https://en.wikipedia.org/wiki/Big_Five_personality_traits)
  + Communication styles: [Psychology Today - Communication Styles](https://www.psychologytoday.com/us/blog/the-main-ingredient/201908/the-4-basic-communication-styles)
* **Approach**:
  + Define 3 persona scripts (Oversharer, Verbose, Reserved)
  + Use semi-scripted prompts to simulate realistic conversations.

### b. ****LLM Deployment****

* **Models**:
  + LLaMA, Mistral, Vicuna, Dolphin — run via [Ollama](https://ollama.com/)
  + Alternatively: [Hugging Face Transformers](https://huggingface.co/models)
* **Dynamic Behavior Tools**:
  + [LangChain](https://www.langchain.com/)
  + Retrieval-Augmented Generation (RAG): [Haystack](https://haystack.deepset.ai/) or LangChain RAG modules

### c. ****Experimentation****

* **Simulation Tasks**:
  + Create chatbot interfaces for interactions.
  + Use controlled prompts per persona type.
* **Log Metrics**:
  + Tone, empathy, coherence, verbosity, context retention

### d. ****Evaluation****

* **Linguistic Analysis Tools**:
  + [NLTK](https://www.nltk.org/), [spaCy](https://spacy.io/)
  + Sentiment & tone detection: [TextBlob](https://textblob.readthedocs.io/en/dev/), [Transformers Sentiment Models](https://huggingface.co/models)
* **Ethical Frameworks**:
  + [FATML Principles](https://www.fatml.org/)
  + [Ethics Guidelines for Trustworthy AI (EU)](https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai)

## 🧰 ****4. Tools and Technologies****

| Category | Tools/Frameworks |
| --- | --- |
| **Frontend** | React (for UI chat simulation) |
| **Backend** | Python, FastAPI |
| **LLM Runtime** | Ollama, Hugging Face Transformers |
| **Data Analysis** | Pandas, NLTK, spaCy, Matplotlib |
| **Persona Simulation** | LangChain, Custom Prompts |
| **Deployment (Optional)** | Local server, Docker, or University HPC |

## 📦 ****5. Deliverables****

| Deliverable | Description |
| --- | --- |
| **Thesis** | A 10–12k word research paper covering methods, results, and risk assessments. |
| **Prototype App** | Chatbot that shifts strategies across user personas. |
| **Strategy Report** | Linguistic and behavioral analysis of LLM adaptation. |
| **Ethics Report** | Critical reflection on fairness, bias, manipulation risks. |

## 🔗 ****Useful Resources****

* **LangChain Docs**: [https://docs.langchain.com](https://docs.langchain.com/)
* **Ollama**: [https://ollama.com](https://ollama.com/)
* **FATML Resources**: <https://www.fatml.org/resources>
* **Hugging Face Models**: <https://huggingface.co/models>
* **AI Ethics Toolkit** (Partnership on AI): <https://www.partnershiponai.org/ai-ethics-toolkit/>

## ✅ ****Next Steps****

1. Set up Ollama or Hugging Face locally.
2. Script persona-specific prompt templates.
3. Build a basic FastAPI backend with chat endpoints.
4. Design React chat interface (optional).
5. Log and analyze conversation responses.