**ADEGuard – AI-Powered Adverse Drug Event Detection and Severity Mapping**

**Domain:** Healthcare **Function:** AI Engineering  
  
**CureviaAI** is a forward-thinking health-tech innovation company dedicated to transforming healthcare.With a strong focus on real-world problems in drug safety, patient care, and clinical decision support, CureviaAI combines AI, machine learning, natural language processing, and biomedical research to deliver actionable insights.

Since the **COVID-19** pandemic outbreak in early 2020, the world has seen an unprecedented scale of mass vaccinations. With this, reports of **Adverse Drug Events (ADEs)** — ranging from mild reactions to severe, life-threatening complications — have risen significantly. These reactions are captured in large-scale, real-world surveillance systems like the [**Vaccine Adverse Event Reporting System (VAERS)**](https://vaers.hhs.gov/about.html), maintained by the [**CDC**](https://www.cdc.gov/) and [**FDA**.](https://www.fda.gov/)

The **VAERS data**, particularly from 2020 to 2025, offers a treasure trove of information — including structured symptom codes and free-text symptom notes. However, these structured entries are limited and miss context like duration, intensity, or location. The free-text notes, while rich, are noisy and lack standardized severity or age-specific patterns — making it difficult for regulators, healthcare providers, and pharma companies to make informed, timely decisions. Despite the data being available, insight is missing.

To address this gap, **Tony Sharma**, the visionary **Chief Innovation Officer** at **CureviaAI**, has initiated a high-priority AI project to build a system that should automatically:

* Extract ADE spans from unstructured symptom text using Named Entity Recognition (NER), supported by structured symptom fields as weak labels
* Cluster symptom variants and uncover age-specific patterns
* Classify ADE severity (mild, moderate, severe) using symptom narratives, structured VAERS fields, and patient metadata
* Provide explainable, audit-ready insights to support regulatory and clinical teams

For this critical mission, Tony turns to **Peter Pandey**, **Lead AI Engineer at CureviaAI**.

**Task:**   
Imagine yourself as **Peter Pandey**. As Peter Pandey, your job is to design and implement a robust AI-powered system, **ADEGuard**,that leverages NLP to detect Adverse Drug Events (ADEs) from free-text symptom descriptions, cluster symptom variants (with modifiers and age-awareness), and classify their severity. 

**Key Requirements:**

* Create Gold Data: Manually annotate a representative sample of narratives with ADE and DRUG spans.
* Use Weak Supervision: Generate weak labels from available report data
* Extract ADE spans (NER): Use BioBERT or similar biomedical NER models for ADE and DRUG entity extraction.
* Modifier-Aware & Age-Specific Clustering: Group ADE symptoms based on modifiers (mild/moderate/severe) and patient age groups.
* Label Severity Levels: Determine severity using rules, manual labels, and BioBERT classification.
* Streamlit UI Visualization: Show token-level highlights, clustering plots, and explainability visualizations (SHAP, LIME).

**Goal:**   
Assist hospitals, regulators, and pharma companies with real-time ADE insights to improve drug safety, optimize response time, and prevent critical complications across age groups.

**Tech Stack**:

1. Python: Core programming language
2. BioBERT/ SciSpacy: NER and embeddings
3. Label Studio/ Prodigy: Annotation
4. HuggingFace Transformers: Classification
5. HDBSCAN + Sentence-BERT: Clustering
6. Streamlit/ Gradio: UI
7. SHAP/ LIME: Explainability

**Note: You're encouraged to use any tools or technologies that enhance your solution**

**Resources Provided**:

* Dataset Information [Click Here](https://vaers.hhs.gov/data/datasets.html?)
* Dataset
* Annotation Guidelines
* Starter GitHub Repo: [Click Here](https://github.com/codebasics/ds-rpc-02)

**Submission Deadline:** 20/09/2025

**Note:**

* We recommend you create a video presentation of 15 minutes or less for the business stakeholders. Additionally, make a LinkedIn post that includes relevant links, your video presentation, and a reflection on your experience while working on this challenge.
* For inspiration, feel free to explore following presentations:
  + [Sample Presentation 1](https://www.linkedin.com/posts/shail-sahu_codebasics-codebasics-resumechallenge-activity-7112519099292745729-tipX/?utm_source=share&utm_medium=member_desktop)
  + [Sample Presentation 2](https://www.linkedin.com/posts/sakshi-chaudhari-a0b989269_codebasics-resumeprojectchallenge-rag-activity-7345839750043394048-A3u1?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEYGhN8BF_joLyeJYuTuPqbiEgTP5VbgB5I)
* Carefully review the evaluation criteria in the “evaluation criteria” document.
* Submit your LinkedIn post link on the Resume Project Challenge page of Codebasics before the deadline. If the post link is not submitted before the deadline, we won't be able to consider it.

**Ready to Step into the Shoes of Peter Pandey?**

You're not just building a model. You're building a system that empowers doctors, supports regulators, and ultimately protects lives.

**The data is public. The need is real. Your mission starts now.**

Wishing you the best of luck from **Team Codebasics!**

**Need help?**

Join the discussion and get support on [**Discord server**](https://discord.gg/bTs2cm2dGA)!