### **Functional Description of the Ensemble Model**

The ensemble model integrates multiple AI and mathematical modeling approaches to address challenges in potato cultivation, particularly those arising from climate change and emerging plant diseases. Its primary goal is to ensure sustainable agricultural practices, optimize crop traits for resilience, and align interventions with Sustainable Development Goals (SDGs).

Find Model Repository here: https://github.com/ShadrackOdikara/FutureProofAgricAIBreedingEpidemiology.git

#### **Key Components:**

**Transformer Models:**

* **Text Generation:** A fine-tuned auto-regressive transformer model ([Llama-3.2-1B-Instruct]) generates insights on potato varietal traits, genetic markers, and socio-economic dynamics. It predicts optimal trait combinations for specific climate scenarios and disease prevalence, aiding breeders and stakeholders in decision-making.
* **Semantic Scoring:** A fine-tuned encoder-only transformer model ([MPNet-based]) classifies the semantic alignment of generated outputs with SDGs. Confidence scores indicate how closely interventions align with SDG targets, aiding in progress assessment.

**Epidemiological Modeling:**

* Simulate disease dynamics, incorporating environmental dependencies and triggers by predicting future climate conditions (temperature, humidity, wind speed) . Then integrates environmental data to predict pathogen proliferation and disease risk. These models factor in climate triggers like temperature, soil pH, and wind speed, creating risk vectors for potato-growing regions. These are represented in a network model with nodes representing regions with their associated risk factors; edges representing connectivity (trade routes, distances). Risk vectors transform as they propagate through the network, creating risk matrices for regions and highlighting disease spread pathways.
* **Retrieval-Augmented Generation System (RAGs):**

Integrates real-time epidemiological data (updated daily via API) into the ensemble model. RAGs allow for dynamic alignment of climate and disease risks with AI-driven recommendations

**Markov Chain Models:**

Used for future state prediction of SDG targets based on current alignment scores and cropland connectivity outputs. The confidence scores from the semantic scoring model serve as states for continuous-time Markov chains to forecast progress toward SDGs..

#### **Data Sources and Methodology:**

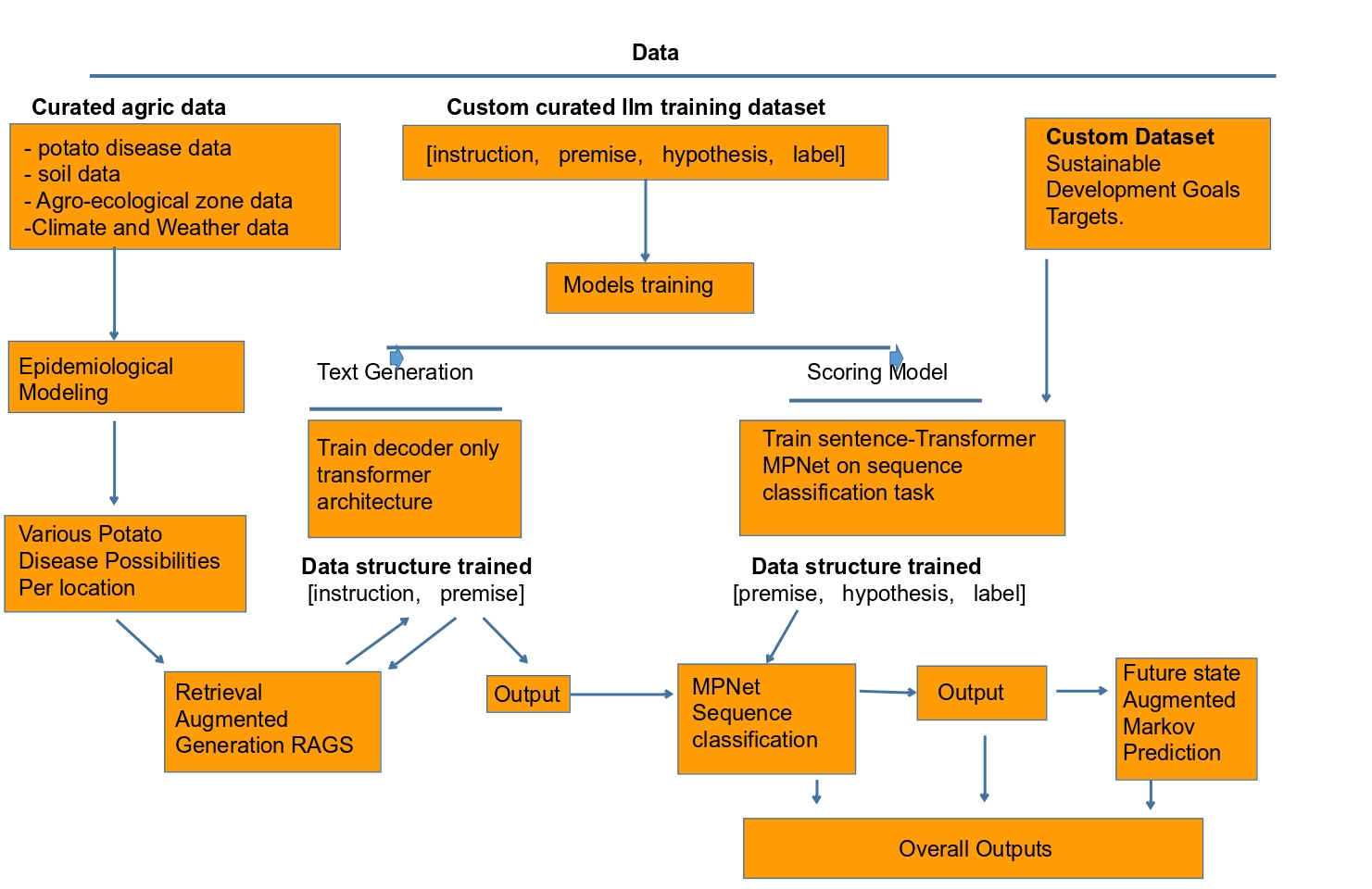
* **Data Sources:** Curated text and tabular data on potato traits, genetic markers, socio-economic dynamics, and climate. Data were sourced from CIP reports, Agricrops databases, Wageningen University, and weatherData.org.
* **Fine-Tuning and Quantization:** Llama-3.2 was fine-tuned on instruction and premise part of the dataset while using Low-Rank Adaptation (LoRA) to reduce computational overhead. The MPNet model was fine-tuned on premise-hypothesis-label part of the datasets to classify semantic relationships.

#### **Output and Applications:**

1. **Disease Risk Assessment:** Provides early warnings for disease outbreaks based on climate conditions and pathogen proliferation risks in potato growing regions of Kenya, Uganda, Peru, China and India.
2. **Varietal Recommendations:** Suggests potato varieties optimized for specific regions based on current and forecasted climatic conditions and the projected disease that is likely to occur by suggesting sustainable resistant varieties.
3. **SDG Alignment:** Evaluates agricultural practices' alignment with SDG 2 (Zero Hunger), SDG 13 (Climate Action) among others using semantic scoring and Markov predictions.
4. **Dynamic Risk Analysis:** Offers region-specific risk matrices for disease management, enabling targeted interventions and policy planning and agro-chemical companies.

#### **Results:**

* The ensemble model demonstrated robust functionality, with fine-tuned Llama achieving a training loss of 1.49 and evaluation loss of 1.52, despite resource constraints.
* It operates effectively in offline mode for localized insights and real-time alignment with climate and epidemiological data.
* The integration of diverse AI and mathematical models facilitates actionable insights for farmers, breeders, and policymakers, addressing global challenges in potato cultivation amidst climate change.

Figure 1: Schema diagram showing the ensemble models functional relationships

By Shadrack Odikara and Team