## **WEEK 7**

## **PLP - AI FOR SOFTWARE ENGINEERING - GROUP 27**

## **Part 1: Theoretical Analysis**

### **QUESTION 1**

### **Explain how Edge AI reduces latency and enhances privacy compared to cloud-based AI. Provide a real-world example (e.g., autonomous drones).**

**Edge AI**, by bringing artificial intelligence computations closer to the data source, significantly reduces **latency** and enhances **privacy** compared to traditional cloud-based AI systems.

In a cloud-based AI system, data must be transmitted from the edge device (like a sensor, camera, or drone) to a centralized cloud server for processing. This journey involves network transmission delays, potential congestion, and the time taken for the cloud server to process the data and send the results back. This round-trip can introduce substantial latency, making real-time applications challenging.

Edge AI, conversely, performs AI inference directly on the device or on a local edge server. This eliminates the need for data to travel to the cloud and back. By processing data locally, the time taken for decision-making and action is drastically cut, leading to near real-time responses.

Cloud-based AI often requires sensitive data to be sent to and stored on remote servers, raising significant privacy concerns. This central repository of data becomes a single point of failure and a potential target for cyberattacks. Furthermore, organizations managing cloud servers have access to the raw data, which can lead to privacy breaches if not handled with the utmost care and in compliance with regulations.

Edge AI enhances privacy by keeping sensitive data localized. Only processed insights or aggregated, anonymized data might be sent to the cloud, or in many cases, no data leaves the device at all. This "data at rest" and "data in transit" principle is significantly improved as the raw, potentially identifiable information never leaves the user's or organization's direct control. This minimizes the risk of unauthorized access, data breaches, and misuse of personal or proprietary information.

Consider **autonomous drones** used for search and rescue operations or infrastructure inspection as a real-world example.

A **cloud-based AI drone** would continuously stream raw footage to a cloud server for analysis. If the drone needed to react instantaneously to an obstacle, the delay caused by data transmission to and from the cloud could lead to collisions. The constant streaming of potentially sensitive visual data also raises privacy concerns.

An **Edge AI drone**, however, has an onboard AI processor and model. As it captures video, the AI model on the drone itself analyzes the footage in real-time. If it detects an anomaly or an obstacle, it can immediately adjust its flight path or trigger an alert without any network delay. From a privacy perspective, the raw video feed never leaves the drone; only detected anomalies (e.g., "object detected at coordinates X, Y") or anonymized metadata might be transmitted to a central command, if at all. This ensures that sensitive visual data remains on the device, enhancing privacy.

In summary, Edge AI empowers autonomous drones with rapid decision-making capabilities and robust privacy safeguards, making them more effective and trustworthy in real-world applications.

### **Q2: Compare Quantum AI and classical AI in solving optimization problems. What industries could benefit most from Quantum AI?**

**Classical AI** in optimization problems often relies on heuristics and approximations. For large and complex optimization problems, the computational resources required can become prohibitively expensive, and finding the **global optimum** (the absolute best solution) is often intractable. Classical algorithms tend to get stuck in **local optima**, which are good solutions but not necessarily the best overall. Examples of such problems include finding the shortest route for a traveling salesperson or optimizing resource allocation.

**Quantum AI**, on the other hand, leverages the principles of quantum mechanics, such as **superposition**, **entanglement**, and **quantum tunneling**, to potentially solve optimization problems more efficiently than classical computers. This means quantum computers could potentially:

* Find the global optimum for problems where classical computers can only find local optima.
* Solve larger and more complex optimization problems that are currently intractable.
* Find solutions significantly faster than classical algorithms for certain problem classes.

The **industries poised to benefit most** from Quantum AI's capabilities in optimization are those dealing with highly complex, multi-variable problems where even marginal improvements in efficiency or resource allocation can yield substantial economic or societal benefits.

1. **Logistics and Supply Chain:** Imagine optimizing global shipping routes, warehouse placement, and last-mile delivery in real-time. This could lead to massive cost savings, reduced carbon emissions, and faster delivery times.
2. **Finance:** Quantum AI could enable more sophisticated portfolio optimization, risk management, fraud detection, and high-frequency trading strategies by analyzing vast amounts of data simultaneously.
3. **Manufacturing and Engineering:** Quantum AI could optimize production schedules, factory layouts, resource allocation in complex manufacturing processes, and even lead to the design of new materials with specific properties by optimizing molecular structures.
4. **Healthcare and Pharmaceuticals:** This includes accelerating drug discovery by optimizing molecular structures for new drugs, personalizing medicine by optimizing treatment plans, and streamlining hospital resource allocation (e.g., bed and staff scheduling).
5. **Energy:** Quantum AI could optimize power grid management, integrate renewable energy sources more efficiently, and enhance resource allocation in smart grids, helping to prevent blackouts and minimize energy loss.
6. **Aerospace and Defense:** This could involve optimizing flight paths, satellite constellations, and complex mission planning, as well as designing more efficient aerospace components.

While Quantum AI is still in its nascent stages, its potential to revolutionize these industries by solving previously intractable optimization problems is immense, promising breakthroughs in efficiency, innovation, and resource utilization.

### **Q3: Discuss the societal impact of Human-AI collaboration in healthcare. How might it transform roles like radiologists or nurses?**

**Human-AI collaboration** in healthcare is poised to bring about a profound societal transformation, enhancing the quality, accessibility, and efficiency of medical services. This synergistic relationship, where AI augments human capabilities rather than replaces them, promises to elevate patient care, reduce healthcare burdens, and empower medical professionals.

The societal impacts include:

* **Improved Diagnostics and Treatment:** AI can analyze vast datasets (medical images, patient records, genomic data) with incredible speed and accuracy, often identifying patterns that humans might miss. This leads to earlier and more precise diagnoses, improving patient outcomes.
* **Enhanced Efficiency and Reduced Costs:** Automating routine tasks and optimizing resource allocation through AI can significantly reduce operational costs, potentially leading to more affordable healthcare services.
* **Personalized Medicine:** AI can analyze an individual's genetic makeup and medical history to tailor treatment plans, moving towards highly personalized care.
* **Increased Accessibility to Healthcare:** AI-powered diagnostic tools and virtual assistants can extend healthcare services to remote or underserved areas.
* **Data-Driven Research and Drug Discovery:** AI can accelerate medical research by identifying drug targets and analyzing clinical trial data, speeding up drug development.
* **Ethical and Equity Considerations:** Integrating AI also raises crucial questions about algorithmic bias, data privacy, and accountability, necessitating careful ethical frameworks.

This collaboration will not eliminate the need for healthcare professionals but rather transform their roles, allowing them to operate at a higher level of cognitive engagement and patient interaction.

**Radiologists** will see a significant shift:

* **Augmented Diagnostic Accuracy:** AI algorithms are already highly proficient at detecting subtle abnormalities in medical images. Radiologists will increasingly **collaborate with AI as a diagnostic co-pilot**, with AI flagging suspicious areas and prioritizing urgent cases.
* **Focus on Complex Cases and Consultation:** With AI handling many routine screenings, radiologists will have more time to focus on complex, ambiguous cases requiring nuanced human judgment and interdisciplinary consultation. Their role will evolve from primary image interpreters to **clinical integrators and consultants**.
* **Workflow Efficiency:** AI can automate tasks like image prioritization and report generation, reducing the administrative burden.
* **Expertise in AI Oversight and Training:** Radiologists will need to understand AI capabilities and limitations, validate AI outputs, and even contribute to the training of AI models.

**Nurses**, at the forefront of patient care, will have their roles enhanced:

* **Enhanced Patient Monitoring and Early Warning Systems:** AI-powered wearables and sensors can continuously monitor patient vital signs and predict potential deterioration. Nurses will utilize these **AI-driven alerts and insights** to intervene proactively.
* **Streamlined Administrative Tasks:** AI can automate mundane administrative tasks such as charting and scheduling, freeing up nurses to spend more time directly with patients.
* **Personalized Care Planning and Education:** AI can help nurses develop personalized care plans and provide tailored educational content for patients. Nurses will become **facilitators of AI-informed patient education and empowerment**.
* **Telehealth and Remote Care:** AI-powered virtual assistants can handle initial patient queries and triage symptoms, extending the reach of nursing care. Nurses will manage and oversee these AI interactions.
* **Focus on Holistic Patient Care:** By offloading repetitive tasks to AI, nurses can dedicate more time to the humanistic aspects of nursing – providing emotional support, building trust, and addressing psychosocial needs that AI cannot fulfill. Their role shifts towards **complex problem-solving, empathetic communication, and holistic patient advocacy.**

In essence, Human-AI collaboration in healthcare marks a paradigm shift from traditional, often reactive, medical practices to a more proactive, personalized, and efficient system. While requiring adaptation and ethical considerations, this synergy promises to elevate healthcare professionals from data handlers to strategic decision-makers and compassionate caregivers, ultimately benefiting society through improved health outcomes and a more sustainable healthcare system.

## **QUESTION 2**

## **Case Study Critique: AI in Smart Cities - AI-IoT for Traffic Management**

The provided article, "How Artificial Intelligence is Transforming Transportation | Smart Mobility, Autonomous Vehicles, and Traffic Management," highlights the profound impact of AI on various aspects of transportation, including smart traffic management.1 While the article focuses broadly on AI's role, its discussion of "AI-Powered Traffic Management" directly addresses the integration of AI with IoT (Internet of Things) for improving urban sustainability.

### **How Integrating AI with IoT Improves Urban Sustainability**

The article implicitly demonstrates how the integration of AI with IoT significantly enhances urban sustainability, primarily through traffic management:

1. **Reduced Congestion and Emissions:** The article mentions how "AI-driven traffic management systems analyze real-time traffic data, monitor congestion levels, and optimize traffic flow using smart signals and predictive analytics."2 This directly relies on IoT sensors (cameras, loop detectors, GPS data from vehicles) to collect the real-time traffic data.3 AI then processes this data to dynamically adjust traffic light timings ("AI-powered traffic lights in Pittsburgh, USA: Reduced travel time by 25 percent").4 By optimizing vehicle flow and reducing idle time, this integration directly minimizes fuel consumption and, consequently, greenhouse gas emissions, contributing significantly to cleaner urban air and reduced carbon footprint.5
2. **Optimized Resource Utilization:** The article highlights how AI "optimizes vehicle flow." This isn't just about speed; it's about making the most efficient use of existing road infrastructure. By using AI to understand traffic patterns and predict future congestion (based on real-time IoT data), cities can avoid costly expansions of road networks.6 This means less concrete, less land consumption, and a more sustainable use of urban space. Furthermore, the ability to minimize emissions contributes to healthier urban environments, reducing the strain on public health resources.

### **Two Challenges in Integrating AI with IoT for Urban Sustainability:**

While the article touches upon challenges broadly, here are two specific challenges relevant to the integration of AI with IoT for urban sustainability:

1. **Data Security and Privacy (as identified in the article):** The article explicitly states, "AI-powered transportation systems are vulnerable to hacking and cyber-attacks" and "AI-driven systems collect and process vast amounts of personal data." In an AI-IoT traffic management system, countless sensors (cameras, vehicle trackers, smart traffic lights) are continuously collecting sensitive data about vehicle movements, potentially identifying individuals, and even monitoring behavior. This creates a massive attack surface for cyber threats. A successful hack could not only compromise privacy but also disrupt critical city infrastructure, leading to chaos (e.g., manipulating traffic signals to cause gridlock or accidents). Ensuring robust encryption, secure communication protocols, and anonymization techniques for such a vast and interconnected data ecosystem is a monumental challenge.
2. **Interoperability and Standardization:** The article mentions various AI applications across different transportation sectors (autonomous vehicles, public transport, logistics). For a truly "smart city" and a comprehensively sustainable urban mobility system, these disparate AI-IoT systems need to communicate and share data seamlessly.7 However, different manufacturers, service providers, and city departments often deploy systems using proprietary technologies and varied data formats.8 Achieving **interoperability** – ensuring these diverse systems can talk to each other – is a significant hurdle. Without common **standards** for data exchange, communication protocols, and AI model interfaces, cities risk creating siloed systems that cannot fully leverage the collective intelligence required for optimal urban sustainability. This fragmentation can hinder holistic traffic management, emergency response coordination, and the ability to extract maximum value from the vast amounts of IoT data for long-term urban planning.

## **PART 2**

Task 1 – Case 1: Biased Hiring Tool – Amazon  
  
Bias Source:  
The bias originated from the training data used by Amazon’s AI hiring tool. The historical data reflected past hiring decisions that were predominantly in favor of male candidates, particularly in technical roles. As a result, the model learned to associate certain gender-linked terms (e.g., “women’s chess club captain”) with lower suitability scores, penalizing female applicants.  
  
Three Fixes:  
1. Re-train with Balanced Data: Use a gender-balanced dataset that reflects equal representation across roles and outcomes.  
2. Remove Gender-Proxies: Eliminate features or terms in resumes that act as indirect gender indicators.  
3. Introduce Fairness Constraints: Implement algorithmic constraints that enforce demographic parity or equal opportunity during model training.  
  
Fairness Evaluation Metrics:  
- Disparate Impact Ratio: Measure if selection rates differ significantly by gender.  
- Equal Opportunity Difference: Check for equal true positive rates across genders.  
- False Positive Rate by Gender: Ensure one gender isn’t penalized more with false rejections.

Task 2 – Case 2: Facial Recognition in Policing  
  
Ethical Risks:  
- Wrongful Arrests: Facial recognition systems have shown higher error rates when identifying individuals from minority groups, increasing the risk of false accusations and unjustified arrests.  
- Privacy Violations: The use of mass surveillance with facial recognition often occurs without individuals' consent, raising major privacy concerns and eroding civil liberties.  
- Bias Reinforcement: If training data underrepresents certain ethnicities or overrepresents them in criminal datasets, it can lead to systemic discrimination and over-policing.  
  
Recommended Policies for Responsible Deployment:  
1. Mandatory Bias Testing Before Deployment: Ensure all facial recognition systems are tested for demographic bias and accuracy across ethnic and gender groups.  
2. Strict Use Regulations: Limit use to clearly defined scenarios (e.g., violent crimes) with judicial oversight.  
3. Transparency and Accountability: Publicly disclose where and how the technology is used, and establish independent oversight bodies for auditing.  
4. Opt-in Consent (where applicable): Use opt-in systems for non-criminal environments like airports or public venues.

## **Task 3**

## **Ethics in Personalized Medicine - Cancer Genomic Atlas**

The Cancer Genomic Atlas (TCGA) is a monumental effort to catalog genetic mutations responsible for cancer, providing invaluable data for personalized medicine.1 However, using AI with TCGA (or similar genomic datasets) to recommend treatments carries significant ethical risks, particularly concerning **bias**.

### **Potential Biases in AI Treatment Recommendations**

The primary bias risk stems from the **underrepresentation of diverse ethnic and ancestral groups** within genomic datasets like TCGA. Historically, genomic research, including TCGA, has been heavily skewed towards individuals of European descent.2 While TCGA has made efforts to increase diversity, a significant imbalance still exists.

This underrepresentation leads to **biased AI models**:

1. **Reduced Efficacy for Underrepresented Groups:** If an AI model is predominantly trained on genomic data from one population group, its ability to accurately identify clinically relevant mutations, predict drug responses, or recommend optimal treatments for individuals from underrepresented ethnic groups (e.g., African, Asian, Hispanic populations) may be significantly compromised. Genetic variations, disease prevalence, and drug metabolism can differ substantially across ancestries.3
2. **Reinforced Health Disparities:** AI recommendations could systematically perform worse or even be harmful for certain populations, effectively embedding and exacerbating existing health disparities in cancer care. This could lead to misdiagnoses, delayed or ineffective treatments, and poorer outcomes for those already facing systemic barriers to healthcare.
3. **Lack of Trust and Adoption:** If patients perceive (or experience) that AI-driven personalized medicine consistently fails or performs poorly for their demographic, it will erode trust in the technology and healthcare system, hindering its adoption and broader societal benefit.

### **Fairness Strategies**

Addressing these biases requires proactive, multi-faceted fairness strategies:

1. **Diverse and Representative Training Data:**
   * **Prioritize Data Collection:** Actively invest in and incentivize genomic sequencing efforts focused on significantly increasing the representation of currently underrepresented ethnic and ancestral groups in datasets like TCGA. This means dedicated funding, community engagement, and ethical data-sharing agreements globally.
   * **Data Augmentation & Synthetic Data (with caution):** Explore techniques to augment existing minority data or generate synthetic data (while rigorously validating its statistical properties and avoiding hallucination) to balance dataset representation, especially when real-world data collection is slow.
2. **Fairness-Aware AI Model Development:**
   * **Bias Detection & Mitigation Algorithms:** Employ specific AI algorithms designed to detect and mitigate bias during model training (e.g., adversarial debiasing, re-weighting data samples).4
   * **Group-Specific Performance Metrics:** Beyond overall accuracy, evaluate model performance (accuracy, recall, precision) independently across different demographic groups to identify and address disparities. Prioritize models that achieve equitable performance.
   * **Explainable AI (XAI):** Implement XAI techniques to understand *why* an AI makes a particular recommendation. This transparency can help identify if the model is relying on spurious correlations or biased features related to ethnicity rather than true biological markers.
3. **Human Oversight and Clinical Validation:**
   * **Clinician-in-the-Loop:** Ensure that AI recommendations are always reviewed and validated by human oncologists and medical professionals who can apply clinical judgment, consider individual patient context, and account for potential AI biases.
   * **Real-World Monitoring:** Continuously monitor the performance of deployed AI systems in diverse clinical settings to detect emerging biases and health disparities in real-world patient outcomes.5

By proactively pursuing these strategies, we can move towards a future where AI in personalized medicine truly benefits all patients equitably, rather than inadvertently widening the health equity gap.

## 

## PART 3 Bias Audit Report – COMPAS Dataset We conducted a fairness audit on the COMPAS Recidivism dataset using IBM's AI Fairness 360 toolkit. The goal was to assess whether the risk scores used to predict recidivism exhibited racial bias, particularly between African-American and Caucasian individuals. Our analysis focused on key fairness metrics including disparate impact, false positive rates, and equal opportunity difference. The results revealed that African-American defendants were significantly more likely to receive higher risk scores than Caucasian defendants, even when controlling for actual recidivism outcomes. The false positive rate (predicting high risk when the person did not reoffend) was higher for African-American individuals, while the false negative rate was higher for Caucasians. This asymmetry indicates a systemic bias in how risk is assessed across racial groups, potentially leading to unfair treatment in sentencing or parole decisions. To visualize this disparity, we plotted risk score distributions and confusion matrices segmented by race. These confirmed the numerical findings and highlighted the need for bias mitigation. Remediation Steps Proposed: 1. Implement pre-processing techniques like reweighting or disparate impact remover to balance the dataset. 2. Use in-processing methods such as adversarial debiasing during model training. 3. Evaluate post-processing techniques to calibrate predictions after training. We recommend that future implementations of such risk assessment tools prioritize transparency, auditability, and fairness metrics monitoring before deployment. Regular ethical reviews and stakeholder engagement are also essential to ensure accountability in criminal justice AI applications.

## **Concept Paper: "Eco-Synapse": An AI-Powered Planetary Health Optimization System for 2030**

**Problem Solved:**

By 2030, global climate change impacts are projected to intensify, leading to more frequent and severe extreme weather events, accelerated biodiversity loss, and increased resource scarcity. Current climate models, while advanced, often struggle with the sheer complexity and interconnectedness of Earth's systems, leading to localized prediction inaccuracies and suboptimal intervention strategies. Furthermore, human decision-making is often slow and fragmented when faced with vast, real-time environmental data.

"Eco-Synapse" aims to address this by providing a hyper-localized, real-time, and predictive AI-powered system that optimizes environmental interventions and resource allocation to enhance planetary health and mitigate climate change impacts. It will move beyond mere monitoring to active, data-driven management of ecological systems.

**AI Workflow:**

**System Name:** Eco-Synapse (Ecological Synaptic Network)

**Core Function:** Real-time planetary health optimization and climate resilience enhancement.

1. **Data Inputs:** Eco-Synapse will ingest a massive, continuous stream of multi-modal data:
   * **Satellite Imagery & Remote Sensing:** High-resolution spectral data, LIDAR, SAR data for land use, deforestation, ice melt, ocean temperatures, biomass, and atmospheric composition (e.g., GHG concentrations).
   * **IoT Sensor Networks:** Ground-based sensors for soil moisture, air quality (particulates, pollutants), water levels, seismic activity, localized weather patterns, and biodiversity markers (e.g., acoustic sensors for wildlife).
   * **Climate Models & Historical Data:** Outputs from existing climate models, decades of historical climate data, and ecological baselines.
   * **Socio-economic Data:** Population density, infrastructure maps, agricultural yields, energy consumption patterns, and human activity data (anonymized where necessary) to understand human-environment interactions.
   * **Genomic and Biological Data:** Data on species health, disease vectors, and ecosystem vitality.
2. **Model Type & Processing:**
   * **Hierarchical Deep Learning Networks:** A combination of Convolutional Neural Networks (CNNs) for image and spatial data analysis, Recurrent Neural Networks (RNNs) for time-series data prediction (e.g., weather patterns, drought cycles), and Graph Neural Networks (GNNs) to model the interconnectedness of ecological systems.
   * **Reinforcement Learning (RL) Agents:** For identifying optimal intervention strategies. RL agents will simulate various climate engineering techniques (e.g., targeted reforestation, precision agriculture, smart water management) and learn which actions yield the best long-term ecological outcomes based on predicted environmental responses.
   * **Digital Twin Simulation:** A high-fidelity digital twin of regional or global ecosystems will be constantly updated with real-time data. This digital twin will serve as a sandbox for the RL agents to test interventions and predict their cascading effects before real-world deployment.
   * **Explainable AI (XAI) Modules:** To provide transparency into the AI's recommendations, allowing human experts to understand the rationale behind proposed interventions and build trust.
3. **Outputs & Interventions:**
   * **Predictive Analytics:** Early warning systems for extreme weather, ecological collapse, resource scarcity.
   * **Optimized Intervention Recommendations:** Specific, localized advice for actions such as:
     + Precision re-wilding and reforestation areas.
     + Targeted water conservation and irrigation strategies.
     + Optimal carbon sequestration locations.
     + Proactive wildfire prevention and suppression strategies.
     + Dynamic resource allocation for disaster response.
   * **Real-time Environmental Dashboards:** Visualizations and alerts for policymakers, environmental agencies, and local communities.

**Societal Risks and Benefits:**

**Benefits:**

* **Enhanced Climate Resilience:** By providing real-time data and predictive insights, Eco-Synapse can enable communities and governments to prepare for and adapt to climate change more effectively, reducing loss of life and property.
* **Optimized Resource Management:** Precision recommendations for agriculture, water use, and energy grids will lead to vastly more efficient resource utilization, ensuring sustainability for growing populations.
* **Biodiversity Preservation:** By identifying at-risk ecosystems and recommending targeted conservation efforts, Eco-Synapse can play a crucial role in halting and reversing biodiversity loss.
* **Improved Human Health:** Reduced pollution, better air quality, and more stable ecosystems directly contribute to better public health outcomes.
* **Accelerated Scientific Discovery:** The vast datasets and advanced models within Eco-Synapse will generate unprecedented insights into Earth's complex systems, accelerating climate science research.

**Risks:**

* **Algorithmic Bias and Inequity:** If the AI models are trained on incomplete or biased data, or if interventions are designed without considering socio-economic factors, Eco-Synapse could inadvertently exacerbate existing inequalities, disproportionately impacting vulnerable communities. For example, interventions might favor economically powerful regions over marginalized ones.
* **Unintended Consequences of Interventions:** Climate engineering or large-scale ecological interventions carry inherent risks of unforeseen negative impacts. While the digital twin aims to mitigate this, complex systems can exhibit emergent behaviors not fully captured in models. A miscalculation or an unpredicted feedback loop could worsen environmental conditions.
* **Centralized Control and Misuse of Power:** The immense power and knowledge derived from Eco-Synapse could be centralized, leading to authoritarian control over environmental policies. A single entity or nation wielding such influence could impose decisions that benefit themselves at the expense of others, or even use the system for surveillance or ecological warfare.
* **Data Security and Privacy Concerns:** The collection of vast amounts of environmental and socio-economic data, even if anonymized, presents significant data security and privacy challenges. Breaches could expose sensitive information or lead to system manipulation.
* **Dependency and "Black Box" Problem:** Over-reliance on the AI system without human understanding of its underlying mechanisms could lead to a "black box" problem where critical decisions are made by an opaque algorithm. This could erode public trust and hinder accountability if errors occur.

Eco-Synapse represents a bold step towards a more sustainably managed planet. Realizing its full potential while mitigating its inherent risks will require robust ethical guidelines, transparent governance, international collaboration, and continuous human oversight to ensure it serves humanity and the planet equitably.

**PART 4: Ethical Reflection**

One of the projects I’m working on is a personalized beauty recommendation system under the Leeknty brand, which aims to use AI to suggest products based on skin tone, concerns, and preferences.

To ensure the project adheres to ethical AI principles, we will prioritize data privacy by collecting only necessary user information with clear consent. The system will be trained on diverse and inclusive datasets to avoid bias related to skin color or ethnicity, and we’ll regularly audit the model for fairness.

Transparency is also a key focus — users will be informed of how recommendations are generated, and we will implement explainable AI techniques to allow them to understand why certain products are suggested. Additionally, a human-in-the-loop approach will be maintained for edge cases, allowing real experts to review and intervene when necessary.

By embedding fairness, accountability, and user empowerment into the design process, we aim to build trust and deliver ethical, user-centered AI in the beauty industry.