🧪 Task 3: Ethics in Personalized Medicine (300 words)

The Cancer Genome Atlas (TCGA) contains over 20,000 primary tumor and matching normal samples across 33 cancer types researchgate.net+1mdpi.com+1pmc.ncbi.nlm.nih.gov+2cancer.gov+2en.wikipedia.org+2. While a landmark resource in precision oncology, it exposes critical **ethical concerns around bias and fairness** in AI-based treatment recommendation systems.

⚠ Identified Biases

1. Racial/Ethnic Representation

TCGA predominantly includes individuals identifying as White (≈ 77%), with Black (12%), Asian (3%), Hispanic (3%), and Indigenous/Pacific Islander (< 0.5%) pmc.ncbi.nlm.nih.gov+1pmc.ncbi.nlm.nih.gov+1. Compared to the broader U.S. cancer population (64% White, 5% Asian, 16% Hispanic), this skews heavily toward White individuals, under-representing other groups . This imbalance limits statistical power to detect actionable mutations in minorities, potentially leading to misrepresentation in diagnostic or therapeutic AI models.

2. Age and Disease Stage

TCGA samples skew toward younger patients and earlier-stage cancers due to selection criteria favoring surgically resectable, first primary tumors nature.com+11pmc.ncbi.nlm.nih.gov+11pmc.ncbi.nlm.nih.gov+11. This excludes older patients and those with advanced/metastatic disease, reducing model applicability across the full clinical spectrum.

3. Clinical Data Gaps

High rates of missing race/ethnicity data (e.g., >10–15% for some cancer types) add uncertainty, weakening subgroup analyses .

☑ Fairness Strategies

1. Enrich Dataset Diversity

Augment training data with genomic and clinical records from under-represented groups—sourcing from community hospitals, multi-ethnic biobanks, and global initiatives like the International Cancer Genome Consortium .

2. Audit Model Equity

Implement bias audits during model development. Evaluate treatment

recommendation accuracy across demographic subgroups (race, age, stage) and enforce parity metrics (e.g., equal false-negative rates).

3. Harness Federated Learning
Use privacy-preserving federated learning to integrate diverse datasets without transferring sensitive patient records.

4. Apply Causal and Statistical Adjustments
Use techniques such as re-weighting, stratified sampling, or PCA-based ancestry correction to reduce confounding by demographic factors .

5. Ensure Explainability and Stakeholder Oversight
Adopt transparent, interpretable models and involve ethicists and patient advocates—especially from marginalized communities—throughout model lifecycle design and deployment.

By acknowledging TCGA's limitations and incorporating robust fairness mechanisms, AI tools in precision oncology can better serve diverse patient populations—minimizing the risk of exacerbating existing health disparities.