Analysis: AI Bias and Fairness in Cancer Treatment Using TCGA Data

The Cancer Genomic Atlas (TCGA) is a landmark dataset used in cancer research and precision medicine, offering genomic profiles across numerous cancer types. While it provides an invaluable resource for training AI systems to recommend treatments, it also presents risks of bias that can lead to inequitable healthcare outcomes.

A major concern is ethnic and demographic underrepresentation in the TCGA dataset. Studies have shown that TCGA data is heavily skewed toward populations of European ancestry, with significantly fewer samples from African, Hispanic, and Asian populations. AI models trained on such imbalanced data may fail to generalize well for underrepresented groups. For instance, predictive accuracy for tumor progression or drug response may be high for White patients but substantially lower for others, leading to biased treatment recommendations and potential disparities in care, and also certain ethnic groups, ages and gender are more prone to certain cancers than others .

Moreover, socioeconomic and environmental factors, which influence cancer development and treatment efficacy, are often absent from the dataset. Ignoring these factors can cause AI systems to oversimplify patient profiles, further disadvantaging already marginalized communities.

Fairness Strategies

To mitigate these biases, developers must adopt fairness-aware AI design strategies:

Diversify Training Data: Expand the dataset by incorporating genomic and clinical data from ethnically and geographically diverse cohorts, ensuring balanced representation during training.

Bias Auditing: Use fairness metrics (e.g., equal opportunity, subgroup accuracy) to audit model performance across different populations and identify disparities.

Domain Adaptation & Transfer Learning: Use techniques that adjust models to perform well on underrepresented subgroups, even with limited data.

Community Engagement: Collaborate with diverse patient populations and clinicians to guide data collection, interpretation, and ethical deployment.

By embedding fairness at every stage of AI development, healthcare systems can reduce bias, improve inclusivity, and ensure that precision oncology benefits all patients equitably.