FutureHealth Perspective
PRESENTED BY THE ARF PHARMA COUNCIL

ARF PHARMA COUNCIL

Maintaining Patient Safety During Every Phase of Agentic AI Usage

Bias in using AI in healthcare is far more than just a data problem because it can **appear at any stage of Agentic AI's development**. Maintaining **constant vigilance from design through deployment** is the only way to **ensure patient safety remains** at the center of innovation.

Phases of AI Implementation

1. Design (Create)

Bias arises from the choices made when conceptualizing the AI tool, such as selecting models, and data sources, defining goals.

2. Develop (Construct)

Bias arises in the training and refining of the algorithm.

3. Deploy (Execute)

Bias occurs once the tool is out in the clinical or pharmaceutical environments being used by humans.

Types of Bias That May Arise

Confirmation Bias

Might develop a "belief" based on initial, skewed training data and favor new data that confirms that initial bias

(e.g., highlight studies that show a drug is effective while ignoring studies that report negative results)

Sampling Bias

Primarily trained on historical patient data, which could potentially exclude certain groups

(e.g., may fail to flag heart attacks in female patients because they were trained on typical male symptoms)

Automation Bias

Over trusting the Agentic AI

(e.g., a radiologist might overlook a visible anomaly on an X-ray because an AI agent noted the scan was "clear," leading to a missed diagnosis)

Guardrails to Mitigate Bias

Balanced Objective Function

intentionally test a model against contradictory or "negative" data to break a confirmation loop

(e.g., select models or algorithms that weigh negative clinical trial results as heavily as positive ones to prevent a "belief" loop that a drug is effective when it isn't)

Synthetic Data

Use specialized generative models to fill data gaps, verified by a secondary AI/clinician to ensure clinical accuracy

(e.g., create female cardiovascular "Digital Twins" and audited AI to bridge male centric data deserts)

Reasoning Prompts

A human must provide a logical justification before accepting/rejecting an AI's output

(e.g., require radiologists to manually highlight AI-flagged zones to verify/dispute the diagnosis)

Designer:
Rebecca Marrero, Primary Research Analyst, People Inc. & ARF Pharma Council Young Pro Officer