Supplemental Table 1. Themes and Tradeoffs Around Fairness for AI in Neuroradiology

Empowering versus Disenfranchising
• AI can enhance diagnosis and care by improving imaging speed and quality, triaging cases, detecting pathology, processing imaging data, making predictions about disease trajectories and outcomes, and decreasing the burden on or even the need for radiologists for some tasks. • AI may exacerbate existing inequities in access and care due to non-representative datasets; insufficient tuning to local demographic, disease prevalence, and imaging parameters; unfair cost functions; limited availability; or inadequacy as a substitute for experienced humans.
Accessibility versus Privacy
• AI development and performance benefit from dataset availability and completeness, including images and demographics, disease characteristics and treatments, and time series data. Public datasets, algorithms, and cloud-based computing spur innovation and promote equal access. • AI processes should safeguard or remove individually identifiable patient information, which might include facial data. Individuals, research groups, health systems, and commercial entities may be reluctant to share personal or proprietary information or algorithms. HIPAA safe harbor rules for de-identification may be insufficient for safeguarding privacy [68].
Personalized versus Standardized
• AI may help tailor predictions and treatments to individuals based on data that are difficult for humans to synthesize, including genetic or imaging biomarkers and could improve quality of care compared to using approaches based on standard of care. • Accounting for individual differences may exacerbate inequality for disadvantaged groups if there is limited availability or diversity of training data.
Automated versus Manual

- AI can be a tool to limit human errors and prejudices in decision making
- AI can provide a false sense of objectivity and fairness and perpetuate systemic racism [24]. Human biases may be incorporated into training data. AI operates on human systems with inherent inequalities. Poorly designed objective functions can lead to unintended outcomes. Human oversight and quality control are needed.

Opportunity versus Opportunism

- AI represents a tremendous opportunity to improve care in neuroradiology and medicine.
- Enthusiasm among clinicians, academics and industry partners for AI solutions should be accompanied by a focus on meaningful clinical translation, transparency, standards of practice, expertise, continuing review of local implementation and performance, and balancing benefits and risks among stakeholders.

Supplemental Table 2. Ethics and Fairness Questions Across the AI Development Cycle

Task definition
What is the potential for automated outcomes to differ by protected classes?
Will an automated process lessen or worsen health care disparities?
What are the consequences of error?
What are the processes for discovering unintended biases?
What are the fairness requirements?
Data Collection
Are identifiable data removed from the dataset?
How are the data collected? Were they obtained with informed or waiver of consent?
What is the source of labeled training data and the potential for labeling bias?
Are training data representative of the population to which the algorithm will be applied?
Are training data augmented by synthetic datasets?
Are demographic data present to permit analysis of fairness with respect to protected classes?
Is the dataset large enough to analyze protected classes and their intersections?
Will data collection be performed transparently and equitably?
Model Definition
What assumptions are made in model optimization?
Is there potential for unfairness in the cost function?
Training and Testing
Will an audit be performed of algorithm performance and fairness?
Are there dataset distribution discrepancies between the training data and deployed environment?
Does the algorithm need to be further tuned to locally representative data using domain adaptation methods?
Deployment and Feedback

How will the algorithm be distributed to the medical community?

Are processes documented and transparent?

What resources are needed for the algorithm to work and for outcomes to be acted on?

How are errors in the algorithms held accountable?