Establishing Global Guidelines for Integrating Artificial Intelligence Systems into Clinical Care

Ayush Sangari, Aditya Sood
Stony Brook School of Medicine

Introduction: In an increasingly digitized world, artificial intelligence (AI) is poised to integrate itself into health care systems across the globe in order to improve clinical care by helping create personalized treatment plans, improving patient diagnoses, and increasing accessible care. Global health policies have already begun implementation of AI-based analysis in the areas of the tuberculosis and malaria epidemics. As AI increasingly enters the clinical sphere, specific guidelines will be critical in maintaining trust of clinical decision-making. The establishment of CONSORT, SPIRIT, and PRISMA guidelines were undeniably instrumental in ensuring transparency and minimum standards of accuracy in randomized clinical trials, interventional trials, and systematic reviews. A similar industry-wide framework must be created and adopted to facilitate the transition of AI into clinical decision-making.

Objectives: This study establishes a framework of system requirements necessary for the successful and ethical adoption of artificial intelligence systems with an emphasis on guidelines for global adoption.

Methods: This study began by surveying existing regulations and guidelines for the use of artificial intelligence in clinical care from prominent regulatory agencies, non-profits, and think-tanks internationally. Four key topics of consideration were identified to guide this analysis: AI Correctness, Workflow Integration, Privacy, and Liability. These existing regulations and guidelines were then analyzed based upon their abilities to address the identified topics of consideration for diverse health care practices and models. New guidelines were also proposed to meet identified gaps between existing frameworks and current global needs.

Results: A new framework for establishing requirements for integrating artificial intelligence systems into health care was developed. In establishing correctness requirements, this framework concludes that artificial intelligence systems must meet regulatory standards for accuracy, be audited for bias, remain frozen (instead of continuously re-training) in government use, and meet criteria for consent form comprehensibility. For workflow integration, this framework formalizes the notion that global health decisions made by policymakers should not solely be based on the recommendation of an artificial intelligence system, finding that artificial intelligence systems must augment - not replace - clinical decisions. It emphasizes a necessity to include basic artificial intelligence education in health policy training, and it supports classifying artificial intelligence systems using the easier-to-interpret and more patient-centered risk-based scheme as opposed to a complexity-based scheme. To address privacy concerns, such systems must use anonymized training datasets that are representative of the populations they will serve, allow countries to own their health data, use an opt-in model to collect patient data, and transparently disclose how patient data is used. Finally, with regards to liability, this framework finds physicians to be liable for patient care decisions even if compliant artificial intelligence systems were consulted, for creators of artificial intelligence systems to be liable if their systems are misleading or dishonest, and for regulators to be liable for mistakes in the authorization process.

Development and discussion of ethical guidelines surrounding the adoption of AI systems is vital

to the effective use of AI in global health policy. This novel framework can serve as the basis for such a guide.