The Healthcare AI Juggernaut: Is PHI Data Governance Possible?

Jenifer Sunrise Winter[1] and Elizabeth Davidson[1]

[1] University of Hawaii at Manoa, Honolulu HI 96822, USA

Abstract. In this commentary we draw on Giddens's metaphor of the relentless scientific progress of modernity as a juggernaut -- a powerful and dangerous force that is difficult to steer [5, p. 151] – to focus attention on the challenges of governing the personal health information (PHI) data essential to advancing AI and machine learning in healthcare. PHI data include a broad array of data from routine transactions to novel data types such as Internet-of-Things medical devices embedded in clinical settings, daily life and even in human bodies. PHI data governance addresses privacy, security, ownership, and use and reuse of health data as well as the underlying values and interests that shape data governance structures [8]. A tacit assumption is that wide-scale sharing of health data will necessarily serve the public good. Given the economic value that may be exploited from PHI data, such assumptions are not fully warranted [6, 8]. We highlight two attributes of AI deep learning that pose novel challenges to PHI data governance: the *scale and scope of data* consumed by deep learning algorithms, and the *opacity of algorithms* in regards to how data are utilized and new data or results are produced. As deep learning algorithms become more sophisticated, PHI data sources are compiled, linked, and used to develop profiles of individuals' behaviors and for predictive health models [1, 4, 9]. Moreover, trace data that are not specifically related to health are also collected and mined to link everyday activities to health status or behaviors[7]. Second, a lack of transparency in corporate-public data sharing arrangements in AI partnerships, and the inherent opacity of deep learning algorithms raise questions about the feasibility of effective PHI data governance. In some cases, opacity may be intentional [2] to protect corporate intellectual property, while in other cases it may be due to the complexity of AI processing. Opacity is problematic for monitoring what PHI data are used and for understanding the purposes and outcomes of data use. Moreover, PHI data governance structures are unlikely to be sufficient to control the combined momentum of AI, deep learning and aggregation of PHI. We focus on two common approaches: *preemptive privacy regulation* and *informed consent*. Effective regulation depends on clear and comprehensive articulation of regulatory requirements as well as the regulating authority's ability to evaluate and monitor compliance. Both are made more difficult by the scale and opacity of AI. The concept of informed consent is itself challenged so that current informed consent practices will not suffice to govern PHI data utilized in AI deep learning ventures. We conclude that critical research is needed to help "envision alternative futures whose very propagation might help them be realized" [5, p. 154], and that IS scholarship can inform, and thereby help actualize, more desirable futures by illuminating the diverse interests, values, and conflicts underlying technology developments [3] arising for the AI healthcare juggernaut

Keywords: Artificial Intelligence, Personal Health Information, Data governance.