

ARTICLE

Monitoring Mental Health: Legal and Ethical Considerations of Using Artificial Intelligence in Psychiatric Wards

Barry Solaiman^{1,2}, Abeer Malik³ and Suhaila Ghuloum^{2,4,5}

¹Hamad Bin Khalifa University (HBKU), College of Law, Qatar

²Weill Cornell Medicine – Qatar

³Hamad Bin Khalifa University (HBKU), Office of Vice President for Research (OVPR), Qatar

⁴Hamad Medical Corporation, Qatar

⁵Doha Institute for Graduate Studies

Corresponding author: Barry Solaiman; Email: barrysol@cantab.net

Abstract

Artificial intelligence (AI) is being tested and deployed in major hospitals to monitor patients, leading to improved health outcomes, lower costs, and time savings. This uptake is in its infancy, with new applications being considered. In this Article, the challenges of deploying AI in mental health wards are examined by reference to AI surveillance systems, suicide prediction and hospital administration. The examination highlights risks surrounding patient privacy, informed consent, and data considerations. Overall, these risks indicate that AI should only be used in a psychiatric ward after careful deliberation, caution, and ongoing reappraisal.

Keywords: Artificial Intelligence; Mental Health; Law; Hospitals; Surveillance; Suicide

1. Introduction

Global health care systems are under overwhelming pressure to meet the exponentially increasing demand for their services. Mental disorders significantly contribute to such increases in demand in the mental health care context.¹ A crucial element of mental health care is acute inpatient care. In order to provide quality and safe care, effective monitoring and management of inpatients in mental health or psychiatric wards is essential.² Since patient monitoring is typically a labor-intensive activity, inpatient psychiatric units have faced distinctive challenges in this area, ranging from slow and inaccurate decision-making to inadequate availability of the right information at the right time.³ The growing demand for mental health care coupled with limited monitoring resources has created opportunities for promising digital solutions such as artificial intelligence (AI) to help address those challenges. The technology extracts novel insights from vast accumulated data about individuals (i.e., big data) to make predictions and recommendations.⁴

¹Philip Moore et al., *Monitoring Patients with Mental Disorders*, 9TH INT'L CONF. ON INNOVATIVE MOBILE AND INTERNET SERV. IN UBIQUITOUS COMPUTING 65, 65 (Jul. 2015).

²The terms “mental health wards” and “psychiatric wards” are used interchangeably in this Article.

³Fatema Mustansir Dawoodbhoy et al., *AI in patient flow: applications of artificial intelligence to improve patient flow in NHS acute mental health inpatient units*, 7 HELIYON, May 2021 at 1, 7, <https://doi.org/10.1016/j.heliyon.2021.e06993> [<https://perma.cc/6KZ8-2ZCN>].

⁴C. Bleas et al., *Artificial intelligence and the future of psychiatry: Qualitative findings from a global physician survey* 6 DIGITAL HEALTH, Jan.-Dec. 2020, at 1, 2, <https://doi.org/10.1177/2055207620968355> [<https://perma.cc/Z9AJ-NMQE>].

This Article explores the practical, legal, and ethical challenges of deploying AI to provide care and operational efficiency in psychiatric units.⁵ While AI systems have not yet gained specific approval for use in mental health wards, research on its use is presently being undertaken in the United States, the United Kingdom, and other countries.⁶ As such, it is a timely opportunity to flag the relevant legal risks distinct to the psychiatric setting before such systems are used in practice.

This exploration is undertaken in three parts. First, there is an examination of the use of AI in hospital wards generally and the associated legal concerns. Second, the distinction between those concerns in psychiatric wards and other wards is outlined. While there is some crossover between the issues, there are distinct factors concerning privacy, informed consent, and data considerations.⁷ Three scenarios are examined to elucidate the challenges in the mental health context: namely, the use of AI-CCTV camera systems to monitor patients, the use of suicide prediction tools, and the use of AI for administration in mental health wards. These scenarios are intended to elucidate the crucial legal considerations; they are not intended to be exhaustive of all potential matters that may arise in a psychiatric ward. Third, the Article recommends developing guidelines that account for the highlighted risks should AI systems eventually be used in psychiatric wards. Overall, significant caution should be taken in implementing AI in psychiatric wards due to the risks posed to patients. While specific use cases may benefit patients, AI should be used only in narrow contexts following robust legal processes.

2. The Use of AI in Hospitals

A critical inquiry is the extent to which there is a distinction between the legal and ethical concerns of using AI in psychiatric wards and other hospital wards. This inquiry is important because mental health is often treated separately from physical health. Many countries have dedicated mental health policies and laws, such as the Mental Health Act 1983 in England and Wales, with provisions on hospital admission. The question is whether such distinctions may result in different requirements for AI used for inpatients in psychiatric wards compared to other wards. The analysis below indicates that there is overlap, but some pertinent additional considerations arise surrounding patient privacy, informed consent and data. To understand those considerations, it is first helpful to consider some existing applications of AI systems in hospitals.

In hospitals, camera-based systems used to track patients' movements now feature built-in AI algorithms designed to "instantly recognize any improper movement based on pre-defined data templates."⁸ Cameras with intelligent sensors can be trained using historical data to spot specific scenarios,⁹ and predict patient behavior, forecast the number of people expected to visit an ER or a waiting room, identify the most frequently used areas of the hospitals and manage patient and visitor traffic around the hospital to improve patient flow.¹⁰

Such AI systems are being deployed in hospitals around the world. For example, researchers at Eindhoven University of Technology developed an intelligent camera that monitors patients around-the-clock and can detect subtle changes in a patient's face or chest. The AI-enabled camera notifies doctors and nurses of possible complications, allowing timely intervention, which can reduce unexpected deaths in

⁵AI systems may be used in the community care and outpatient context, but the focus of this Article is specially on inpatients.

⁶This Article primarily builds upon literature produced in United States and United Kingdom, but research from other countries is also examined. While the specific laws may differ in countries, the underlying legal risks and solutions are similar.

⁷Similar concerns have been highlighted by the OHCHR in mental health care for the use of smart pills and remote monitoring. See WHO/OHCHR, *infra* note 48, at 42.

⁸S. Balaji, *How cameras and Artificial Intelligence are changing the way we look at patient care?*, E-CON SYSTEMS (Feb. 14, 2022), <https://www.e-consystems.com/blog/camera/applications/how-cameras-and-artificial-intelligence-are-changing-the-way-we-look-at-patient-care/> [https://perma.cc/7XEH-BPWU].

⁹Neal Lorenzi, *Video security systems add variety of features*, HFM MAGAZINE (Apr. 30, 2021), <https://www.hfmmagazine.com/articles/4173-video-security-systems-add-variety-of-features> [https://perma.cc/TU6V-ZFJH].

¹⁰Enav Perez, *Enhancing Video Surveillance at Hospitals with Video Content Analytics*, BRIEFCAM (May 27, 2020), <https://www.briefcam.com/resources/blog/enhancing-video-surveillance-at-hospitals-with-video-content-analytics/> [https://perma.cc/5MXU-Z8Q4].

the hospital.¹¹ A similar application was used in a hospital in Nebraska where AI-enabled cameras were installed across beds for patients deemed at risk of leaving their beds unattended and falling. Around 200,000 hours of video recording data were used to train the AI algorithm to predict when a patient is about to leave the bed and alert the staff on duty if intervention is required.¹²

AI systems are also enhancing operational efficiency and improving security around hospitals. Cameras with facial recognition technology can “identify and authenticate individuals to control access and prevent abduction.”¹³ Great Ormond Street Hospital (GOSH) in the United Kingdom partnered with technology company Arm to develop smart cameras that not only predict a person’s behavior but can also detect intruders by cross-checking a given individual’s physical features with a database of known faces and alert security. The AI system is also expected to assist doctors during surgeries in the future by monitoring equipment usage and ensuring surgical tools are not accidentally left inside patients. The system does not store any live video information or pictures, thus protecting patients’ identities.¹⁴ AI personal assistants may also aid patient monitoring when medical staff are unavailable.¹⁵ Indeed, AI digital assistant monitors have been used at the Cleveland Clinic to monitor 100 beds in 6 ICUs. These systems identify at-risk patients through predictive and advanced analytics by monitoring and learning from conversations between doctors and patients in hospital rooms.¹⁶

Furthermore, AI can predict workloads in emergency departments¹⁷ and match competencies of health care providers to specific tasks for efficient staff rota planning.¹⁸ Thus, AI systems capture video data from physical spaces where sensitive patient care activities take place and interpret the behaviors of health care professionals and patients, to monitor deviations in intended bedside practices, patient mobilisation activities and hygiene,¹⁹ as well as to conduct bed occupancy checks, patients’ vitals monitoring, and patient fall or violence detection, among others.²⁰

2.1 The Legal Concerns of Using AI in Hospitals

While continuous AI monitoring reduces the need for direct patient observation and frees hospital staff from administrative tasks, it is not without challenges.²¹ Gerke *et al.* have mapped out the ethical and

¹¹Smart camera reduces unexpected complications in hospitals, TU/E EINDHOVEN UNIVERSITY OF TECHNOLOGY (Mar. 15, 2021), <https://www.tue.nl/en/news-and-events/news-overview/01-01-1970-smart-camera-reduces-unexpected-complications-in-hospitals/#top> [https://perma.cc/A4RN-XMQL]; see also Serena Yeung *et al.*, *A computer vision system for deep learning-based detection of patient mobilization activities in the ICU*, 2 NPJ DIGIT. MED. (2019) [hereinafter Yeung, *A computer vision system*].

¹²Jessica Kim Cohen, *Nebraska hospital deploys AI, depth cameras to help curb patient falls*, MODERN HEALTHCARE (Jan. 25, 2020), <https://www.modernhealthcare.com/patients/ai-watches-inpatients-bryan-medical-center-help-curb-falls> [https://perma.cc/7BLW-HRYW].

¹³Lorenzi, *supra* note 9.

¹⁴Joseph Archer & Charlotte Krol, *NHS hospital is developing AI cameras that can spot intruders and monitor staff*, THE TELEGRAPH (Oct. 10, 2018), <https://www.telegraph.co.uk/technology/2018/10/10/nhs-hospital-developing-ai-cameras-can-spot-intruders-monitor/> [https://perma.cc/Z65L-EKJ6].

¹⁵DonHee Lee & Seong No Yoon, *Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges*, 18 INT’L J. ENVTL. RES. PUB. HEALTH, Jan. 2021, at 1, 4, <https://doi.org/10.3390/ijerph18010271> [https://perma.cc/BWM2-WCJR].

¹⁶*Id.* at 7.

¹⁷Nan Liu *et al.*, *Artificial intelligence in emergency medicine*, 2 J. EMERGENCY CRITICAL CARE MED., Oct. 2018, at 1, 3, <http://doi.org/10.21037/jeccm.2018.10.08> [https://perma.cc/C82V-LEYJ].

¹⁸Anu Thomas, *How Hospitals Can Tap AI To Manage Staff Better Amid Covid-19 Crisis*, ANALYTICS INDIA MAGAZINE (Apr. 11, 2020), <https://analyticsindiamag.com/how-hospitals-can-tap-ai-to-manage-staff-better-amid-covid-19-crisis/> [https://perma.cc/R9CD-UN4E].

¹⁹Sara Gerke *et al.*, *Ethical and Legal Aspects of Ambient Intelligence in Hospitals*, 323 J. AM. MED. ASS’N 601, 601 (2020).

²⁰Intelligent Video Analytics in Healthcare Industry, MEDIUM (Oct. 11, 2021), <https://medium.com/@Staqui/intelligent-video-analytics-in-healthcare-industry-6b4dae51ecc2> [https://perma.cc/9LEN-K3N9]; Chris Noon, *HealthManagement, The English Patients: This UK Hospital Is Harnessing AI To Deliver Slicker Service*, 20 HEALTHMANAGEMENT.ORG J. 149 (2020).

²¹Yeung, *A computer vision system*, *supra* note 11, at 3.

legal concerns that broadly arise in AI and health care.²² Ethical concerns include informed consent to use, safety and transparency, algorithmic fairness and bias, and data privacy. Legal concerns include safety and effectiveness, liability, data protection and privacy, cybersecurity, and intellectual property.²³ Gerke et al. narrow those considerations when examining aspects of ambient intelligence in hospitals to privacy and reidentification risks, consent and liability.²⁴ To that, we can add data accuracy concerns.

On privacy and reidentification, the risk is that health care professionals and patients may be identified, which could undermine their privacy. While captured data may be deidentified in one data set (for example, by capturing silhouette images instead of color images), data triangulation makes it possible to reidentify those individuals from other data sets (such as doctor shift rota data).²⁵ Additionally, data accuracy risks are associated with smart surveillance due to insufficient data available to train AI algorithms. Developing these algorithms is a cognitively laborious task, which requires obtaining sufficient training examples of scant activities undertaken by the staff and patients around the hospital and manually annotating data from video recordings, which is susceptible to human labelling error.²⁶ Data bias is also a potential pitfall, with some studies finding that algorithms demonstrate significant racial bias when predicting aspects of care.²⁷

Data accuracy is also affected by practical considerations, such as the camera placement in a hospital ward. The location of smart cameras determines an AI-enabled system's effectiveness in capturing and processing accurate real-time data.²⁸ In this regard, two main concerns arise. Firstly, the architectural design of different hospital units does not allow the hospital to install a network of sensors with sufficient overlapping fields of view, making data association and trajectory prediction across all hospital units a cumbersome task. Secondly, limited visual data affects accuracy because important visual appearance cues are lost as de-identified depth images are used instead of color videos due to privacy concerns surrounding personal health information, making tracking personnel extremely difficult.²⁹

These accuracy risks could feed into medical liability concerns. For example, AI systems could make incorrect recommendations that may ultimately harm patients, raising questions about where responsibility lies.³⁰ Patients may also require access to data recorded from an AI system for a liability claim, which raises questions about the appropriate rules a hospital implements for retaining and deleting data. Hospital workers could fail to act on a problem detected by an AI system, which could leave the hospital vulnerable to liability claims. AI systems may also record illegal activities, which requires clear policies about the circumstances in which hospitals will retain and share the recordings with law enforcement.³¹

On consent, the concern is whether explicit consent should be required from patients for AI systems recording them.³² Hospitals have been sued for recording and retaining videos of patients in vulnerable situations in U.S. hospitals.³³ Nevertheless, hospitals might escape liability using consent documents containing boilerplate language, which provide authorization for hospitals to record patients for quality

²²Sara Gerke et al., *Ethical and legal challenges of artificial intelligence-driven healthcare*, in *ARTIFICIAL INTELLIGENCE IN HEALTHCARE* 295 (Adam Bohr & Kaveh Memarzadeh eds., 2020).

²³*Id.*

²⁴Gerke et al., *supra* note 19.

²⁵Gerke et al., *supra* note 22, at 317-18; W Nicholson Price II & I Glenn Cohen, *Privacy in the Age of Medical Big Data*, 25 NAT. MED. 37, 39 (2019).

²⁶Yeung, *A computer vision system*, *supra* note 11, at 11.

²⁷Ziad Obermeyer and others., *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 SCI. 447 (2019).

²⁸Balaji, *supra* note 8.

²⁹Albert Haque et al., *Towards Vision-Based Smart Hospitals: A System for Tracking and Monitoring Hand Hygiene Compliance*, 68 JMLR 75, 77 (2017).

³⁰W. Nicholson Price II et al., *Liability for Use of Artificial Intelligence in Medicine*, in *RESEARCH HANDBOOK ON HEALTH, AI AND THE LAW* (Barry Solaiman & I Glenn Cohen eds., forthcoming 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=41115538 [<https://perma.cc/V5NV-JBVE>].

³¹Gerke et al., *supra* note 19, at 602.

³²*Id.* at 601.

³³*Id.*

improvement and security reasons.³⁴ It has been noted that the more explicit and specific the consent is, the better.³⁵ For hospital employees, the employer has a “strong claim of a right to record their employee without meaningful consent” owing to legitimate business interest and public interest justifications under U.S. law.³⁶ That right to record the employee will be satisfied as long as the recording is taken in a quasi-public space such as an operating room, corridor or patient room, where the goal is to improve patient care, where workers are informed about the recording, and where proper safeguards are in place to protect individual privacy.³⁷ As such, some literature has examined the legal concerns arising in the context of AI being used in hospitals. The query here is the extent to which there is a distinction in the mental health inpatient context.

3. The ‘Siloed’ Considerations in Mental Health

Examining AI systems specific to the inpatient mental health context is challenging. Neither the FDA nor any other regulatory body has approved such devices for use in psychiatric wards (as of 2023).³⁸ One useful starting point is to consider the rules that apply to research on mental health because significant research is being undertaken involving psychiatric inpatients with a view to deploying the technology in the future. Taking a step back, we can map out important international norms that bind such research. When developing AI systems, developers will train those systems on data. Researchers can obtain real-time data from individuals on-site at a hospital or off-site in naturalistic settings such as the home.³⁹ User consent for data collection is required, as are the expected Institutional Review Board (IRB) approvals.⁴⁰

Mental health data is often regarded as more sensitive than other health data; it is, therefore, covered by the Helsinki Declaration on protecting vulnerable patient groups, such as psychiatric patients.⁴¹ Medical research with a vulnerable group is only justified where it is responsive to the health needs of that group, and “should stand to benefit” the group.⁴² The phrasing of “should stand to benefit” amends the previous threshold that such research is justified where there is a “reasonable likelihood” that the population stands to benefit.⁴³ This change broadened the justificatory scope for research involving vulnerable populations such as psychiatric patients. Informed consent is critical in this context. Consent must be voluntary or, if not possible, given by a legally authorized representative.⁴⁴ Additionally, there are well-established principles of the United Nations on protecting persons with mental illness and improving mental care that are relevant to research.⁴⁵ Every patient in a mental health facility shall have

³⁴*Id.*

³⁵*Id.*

³⁶*Id.*

³⁷*Id.* at 602.

³⁸Ellen E. Lee et al., *Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom*, 6 BIOLOGY PSYCHIATRY: COGNITIVE NEUROSCI. NEUROIMAGING 856, 856-57 (2021); for an overview of wearable devices used in mental health, see, Arfan Ahmed et al., *Wearable Devices for Anxiety and Depression: A Scoping Review*, 3 COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE UPDATE (2023), <https://doi.org/10.1016/j.cmpbup.2023.100095> [<https://perma.cc/U652-GL98>].

³⁹Enrique Garcia-Ceja et al., *Mental Health Monitoring with Multimodal Sensing and Machine Learning: A Survey*, 51 PERSVASIVE AND MOBILE COMPUTING, Dec. 2018 at 1, 13, <https://doi.org/10.1016/j.pmcj.2018.09.003> [<https://perma.cc/G335-GSWW>].

⁴⁰*Id.*

⁴¹*Id.* at 14.

⁴²WORLD MED. ASS'N, WMA DECLARATION OF HELSINKI – ETHICAL PRINCIPLES FOR MEDICAL RESEARCH INVOLVING HUMAN SUBJECTS art. 20 (June 1964), <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/> [<https://perma.cc/C6DS-KL7A>].

⁴³Aisha Y Malik & Charles Foster, *The Revised Declaration of Helsinki: Cosmetic or Real Change?*, 109 J. ROYAL SOC'Y MED. 184, 187 (2016).

⁴⁴WMA DECLARATION OF HELSINKI, *supra* note 42, at art. 25-29.

⁴⁵G.A. Res. 46/119, Principles for the protection of persons with mental illness and the improvement of mental health care (Dec. 17, 1991).

the right to privacy.⁴⁶ Clinical trials and experimental treatments shall never be carried out on any patient without informed consent unless approval is given from a competent, independent review body specifically constituted for that purpose.⁴⁷

Thus, from a research perspective, mental health patients are treated with specific norms and considerations in mind. There is also a legal distinction between the rules covering psychiatric patients and other patients. In mental health, specific rules govern diagnostic considerations, voluntary and involuntary admission and treatment, informed consent for special treatments such as electroconvulsive therapy, monitoring and review mechanisms, and the governance and administration of health care services, among others.⁴⁸ Other international norms overarch laws on detention, such as Article 8(1) of the European Convention on Human Rights (ECHR) on the right to respect for private, family life, home and correspondence, which creates an obligation on staff in psychiatric wards to balance the right to privacy of patients with safety.⁴⁹ There is also Article 3 on the right to be free from torture and inhuman and degrading treatment and Article 14 on the right to be free from discrimination.

The “siloed” legal approach in mental health has been criticized for emphasizing segregation, potentially reinforcing stigma, and creating exceptions to the equal exercise of patient rights through arbitrary restrictions and the right to free and informed consent.⁵⁰ Nevertheless, the legal distinction exists. Further, while informed consent is required for any patient in a hospital partaking in such research, the law and guidelines deem psychiatric patients more vulnerable, requiring more checks and balances. The query is how these general principles governing mental health research, and the specific legal provisions in the mental health space, intersect with the legal considerations for AI systems. Below, these considerations are elucidated through an examination of three applications of AI systems in the psychiatric inpatient setting: namely, AI surveillance, suicide prediction, and hospital administration. These applications are offered as illustrative case studies for the analyses in this Article and are not intended to be comprehensive of all potential applications.

3.1 AI-CCTV Surveillance in Hospitals

The manner in which surveillance is implemented is very important to the patient experience. CCTV cameras are a crucial tool for monitoring communal areas shared by patients and staff and seclusion rooms. CCTV may, in certain circumstances, be used in private areas, including patient bedrooms.⁵¹ Continuous CCTV monitoring enables better risk detection in hospital wards, allows patient observations to occur without disturbance, and provides data for training purposes.⁵² However, surveillance “within hospital wards is not based on reciprocity” as it subjects patients to constant monitoring without their consent.⁵³ Hospitals have adopted numerous policies and guidelines generally addressing surveillance, but additional, more specific considerations on privacy and informed consent apply in psychiatric

⁴⁶*Id.* princ. 13(1)(b).

⁴⁷*Id.* princ. 11(15).

⁴⁸WORLD HEALTH ORGANIZATION/OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR HUMAN RIGHTS [WHO/OHCHR], MENTAL HEALTH, HUMAN RIGHTS, AND LEGISLATION: GUIDANCE AND PRACTICE, 12 (2023), <https://iris.who.int/bitstream/handle/10665/373126/9789240080737-eng.pdf?sequence=1> [<https://perma.cc/8GBW-9DP5>].

⁴⁹DEPARTMENT OF HEALTH (UK), MENTAL HEALTH ACT 1983: CODE OF PRACTICE, 63 (2015), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/435512/MHA_Code_of_Practice.PDF.

⁵⁰*Id.* at 28.

⁵¹Yahel E. Appenzeller et al., *Ethical and Practical Issues in Video Surveillance of Psychiatric Units*, 71 PSYCH. SERV. 480, 480 (2020).

⁵²Suki Desai, *The new stars of CCTV: what is the purpose of monitoring patients in communal areas of psychiatric hospital wards, bedrooms and seclusion rooms?*, 6 DIVERSITY IN HEALTH AND CARE at 1, 3 (2009) [hereinafter Desai, *The new stars of CCTV*]; Serena Yeung et al., *Bedside computer vision—moving artificial intelligence from driver assistance to patient safety*, 378 N ENGL. J. MED. 1271, 1273 (2018) [hereinafter Yeung, *Bedside computer vision*].

⁵³Suki Desai, *Violence and surveillance in mental health wards*, CTR. CRIME AND JUST. STUD. 4 (2011), <https://www.crimeandjustice.org.uk/sites/crimeandjustice.org.uk/files/09627251.2011.550146.pdf> [hereinafter Desai, *Violence and surveillance in mental health wards*].

wards, suggesting that much caution should be exerted when monitoring psychiatric patients with AI.⁵⁴ This phenomenon is easily observed when examining the rubric of rules that have developed in England. There, the Home Office has developed a Surveillance Camera Code of Practice that emphasizes CCTV should only be used in the pursuit of a legitimate aim, be necessary, take account of an individual's privacy and be subject to regular review, among other rules.⁵⁵ In the health sphere, the Care Quality Commission (CQC) and Information Commissioner's Office (ICO) provide guidance and regulatory oversight for care in England. The ICO can investigate and take enforcement action where surveillance breaches a patient's rights under the Data Protection Act (DPA) 2018 and U.K. GDPR. To assist organizations, it has developed detailed guidance on the use of CCTV and surveillance.⁵⁶

Privacy and dignity must always be maintained with surveillance.⁵⁷ Cameras can be installed in public and communal areas, but not private areas such as bathrooms. However, covert recordings are permissible if the benefits outweigh other considerations on privacy and dignity.⁵⁸ All recordings must be logged and traceable and not retained for more than 28 days.⁵⁹ Disclosure of recordings can only be made in strict adherence to the DPA 2018 and U.K. GDPR.⁶⁰ Where an individual cannot consent to the use of CCTV, the clinical leads must decide according to the Mental Capacity Act (MCA) 2005.⁶¹ Thus, if CCTV is to be used in a patient's bedroom (especially where a patient lacks mental capacity), permission may be required from the Court of Protection on an ad-hoc basis.⁶²

Also of import is the ECHR. Article 8 articulates a qualified right, which means that recordings can be used despite breaching privacy where the recording is lawful, for a legitimate aim, and proportionate.⁶³ The least restrictive possible alternatives to CCTV must be considered, such as using call bells or pendants, using sensors that detect movements, improving staff training and supervision, or increasing staff numbers.⁶⁴ Depending on the system used, these considerations may be a double-edged sword for AI. CCTV incorporating AI technology may be less likely to gain approval, but an AI-enabled wearable may be more likely to gain approval because it may be deemed less restrictive.

The British Institute of Human Rights (BIHR) provide case studies for these considerations. They discuss a man with autism who sometimes harmed himself. Staff installed a CCTV camera in his room because that was deemed less restrictive than posting a staff member in his room twenty-four hours a day; they determined that would be a legitimate and proportionate response to the circumstances. The staff then installed CCTV in all the rooms of the facility, but they were not switched on. The courts considered the matter and decided that CCTV could be used in the man's bedroom, but its use was deemed a significant restriction of his Article 8 ECHR rights, requiring that a judge regularly review the use of the CCTV camera in his bedroom. It was held that using CCTV in other bedrooms was not lawful, legitimate, or proportionate, so they were removed.⁶⁵

⁵⁴Desai, *The new stars of CCTV*, *supra* note 52; See Yeung, *Beside computer vision*, *supra* note 52.

⁵⁵HOME OFFICE (UK), SURVEILLANCE CAMERA CODE OF PRACTICE, 9 (June 2013), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1035067/Surveillance_Camera_CoP_Accessible_PDF.pdf

⁵⁶INFORMATION COMMISSIONER'S OFFICE (ICO), GUIDANCE ON VIDEO SURVEILLANCE INCLUDING CCTV (Oct. 14, 2022), <https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-video-surveillance-including-cctv-1-0.pdf>

⁵⁷SOLENT NHS TRUST, POLICY FOR SURVEILLANCE CAMERA SYSTEM, 6 (2017), <https://www.solent.nhs.uk/media/2075/ig08-surveillance-camera-system-policy-v3.pdf>.

⁵⁸*Id.* at 7.

⁵⁹*Id.* at 9.

⁶⁰*Id.* at 10.

⁶¹*Id.* at 6.

⁶²Care Quality Commission, *Using Surveillance: Information for Providers*, at 11 (2008). The Court of Protection is a special court in England created under the MCA 2005 to make decisions for those lacking capacity on financial or welfare matters.

⁶³THE BRITISH INSTITUTE OF HUMAN RIGHTS, HUMAN RIGHTS AND THE USE OF CAMERAS AND OTHER RECORDING EQUIPMENT IN HEALTH & SOCIAL CARE: A SHORT GUIDE, 2 (2021).

⁶⁴*Id.* at 3.

⁶⁵*Id.* at 4.

Such examples raise questions about the appropriate limits of surveillance. It is wrong to assume that psychiatric patients require higher levels of supervision. Applying blanket surveillance rules without considering the impact on individuals could cause more harm and distress than help.⁶⁶ This was the case for another patient detained under the law and subject to CCTV surveillance installed in her room and every other room in addition to “arm’s length” observations by two staff day and night. The patient found the CCTV to be intrusive, inhumane and degrading.⁶⁷ Consequently, surveillance must be carefully planned, considering what is legitimate, necessary, and proportionate. A major trade union has recommended to members that privacy impact assessments be conducted.⁶⁸ The CQC has noted that it is up to the providers to decide whether to use surveillance (within the limits of the law), but that “we would be concerned by an over-reliance on surveillance to deliver key elements of care, and it can never be a substitute for trained and well supported staff.”⁶⁹

While such legal protections exist in England, their implementation can be problematic; privacy can be quite limited. The Mental Health Act Revised Code of Practice places an obligation on public authorities to respect a person’s right to a private life, including those detained under the Act.⁷⁰ However, CCTV monitoring of communal areas used by patients—not all of whom have consented to such surveillance—raises privacy concerns. The right to privacy is further curtailed for involuntarily detained psychiatric patients. Since their access to space outside the ward environment is restricted due to the involuntary nature of admission, such patients have limited areas where they can avoid constant monitoring.⁷¹ These risks may be replicated when AI is incorporated into such technologies. The risk of constant surveillance remains and may be heightened if the form of monitoring encompasses wearable devices that constantly track a patient, no matter their location in a ward.

Despite these concerns, the research in this area is not clear-cut. One study in Australia notes that CCTV cameras do not necessarily upset patients or motivate them to spend time in locations without camera surveillance.⁷² Some patients may be passive about CCTV monitoring, which could make them feel safe, particularly on open wards.⁷³ The views of staff should also be considered. Many staff in psychiatric wards have been victims of assault by patients, and surveillance could protect them.⁷⁴ Staff are also concerned about how to use AI systems to identify such risks.⁷⁵ The key is to implement security measures that do not provoke aggression or violence. In this regard, the characteristics of the ward are essential, and due consideration should be given to how AI systems could and should fit within that environment. The CQC has noted that many psychiatric units are in old and outdated buildings, which can lead to issues around privacy and dignity for patients. Wards can be “barren, visually impoverished environments dominated by security fencing” that can undermine morale and have a detrimental impact on recovery.⁷⁶ The same is true of most psychiatric wards in the world, which are in old buildings that do

⁶⁶*Id.* at 7.

⁶⁷*Id.* at 6.

⁶⁸UNISON, *USE OF SURVEILLANCE IN HEALTH AND CARE SETTINGS: GUIDANCE FOR UNISON REPRESENTATIVES*, 2 (2015), <https://www.unison.org.uk/content/uploads/2015/02/TowebUNISON-guidance-on-the-use-of-surveillance-cameras-in-health-and-care-settings2.pdf>.

⁶⁹*Id.* at 4; see also *Care Homes: CCTV*, *Parliamentary Debates*, HOUSE OF COMMONS (Sept. 5 2018) <https://hansard.parliament.uk/commons/2018-09-05/debates/B518C2CC-DA1E-444A-88F5-505FBEFFCE3C/CareHomesCCTV> [https://perma.cc/W29U-EFYD].

⁷⁰Desai, *The new stars of CCTV*, *supra* note 52.

⁷¹*Id.*

⁷²Clemence Due et al., *Surveillance, Security and Violence in a Mental Health Ward: An ethnographic case-study of an Australian purpose-built unit*, 10 *SURVEILLANCE & Soc’y* 292, 301 (2012), <https://pdfs.semanticscholar.org/d281/259a5d5237557f2adb7ef016ae79cecb2f2b.pdf>.

⁷³*Id.* at 296 & 301.

⁷⁴*Id.* at 301.

⁷⁵See Ben Greer et al., *Predicting Inpatient Aggression in Forensic Services Using Remote Monitoring Technology: Qualitative Study of Staff Perspectives*, 21 *J. MED. INT. RES.* Sep. 2019, at 1, <https://www.jmir.org/2019/9/e15620> [https://perma.cc/2CVT-2R6K].

⁷⁶Care Quality Commission, *Monitoring the Mental Health Act in 2021/22*, at 46 (2022), <https://www.cqc.org.uk/publications/monitoring-mental-health-act/2021-2022/ward-environments> [https://perma.cc/QMT2-DRYS].

not allow for proper CCTV surveillance. The use of AI tracking systems in those environments should, therefore, only be done with careful planning.

Another challenge is gaining informed consent for using CCTV cameras due to psychiatric patients' varying levels of perception and cognition which can result in an impaired or fluctuating capacity to consent. Certain conditions, such as those linked to feelings of paranoia, tend to incite violent reactions towards the existence of cameras.⁷⁷ While it is good practice under the MCA to obtain informed consent for CCTV monitoring, how and when such consent should be sought is ambiguous. Although there is a general presumption of surveillance in public spaces at hospitals not requiring consent, the same cannot be said when CCTV is used to monitor communal areas, seclusion rooms or patient bedrooms. These risks are deepened and broadened where AI is used because of the more pervasive and detailed information that such systems may potentially generate. Even if there is consent for using an AI system in a specific way, the autonomous nature of the system could result in recommendations and analyses for which consent was not obtained initially.

Thus, it is challenging to determine where the balance should tilt on using AI-enabled surveillance systems. There are obvious risks and strict legal considerations for the use of CCTV generally, but should those risks entirely preclude their use for psychiatric inpatients? AI-enabled surveillance does not yet exist in mental health wards, but one could consider prison surveillance by analogy. AI surveillance is used on prisoners to detect and prevent harmful incidences in China, Hong Kong, the United Kingdom, the United States, India and South Korea.⁷⁸ Some systems can detect changes in parameters up to thirty minutes before an observable act of aggression with greater precision than existing methods.⁷⁹ In Hong Kong, AI surveillance systems can detect instances of self-harm and alert staff, and detect if a person has any contraband.⁸⁰ Wearable wristbands can track vital signs and flag any signs of injury. Robots in South Korea can monitor violence and suicide risks.⁸¹ In the United States, mass-monitoring millions of phone calls through AI has helped officials combat violence, drug smuggling and attempted suicides in near real-time. By searching for keywords, phrases, and prison slang, investigators notify law enforcement officers when the AI system picks up suspicious language. This has prevented dozens of attempted suicides by providing inmates with psychological counselling in the minutes and hours after the inmate was recorded making references to harming themselves, threatening violence, or plotting to smuggle in contraband.⁸²

Of course, the analogy cannot be stretched too far. Inpatients are not inmates in prisons, and one should not generalize the characteristics of psychiatric inpatients to be at risk of suicide or harm to others. Some inpatients are voluntarily there. Nevertheless, there are some similar paradigms and risks that exist that are helpful to consider, such as patients being involuntarily detained, those at risk of self-harm or suicide, those possessing contraband that can be a risk to themselves or others, and those posing a risk of violence (a common occurrence on psychiatric wards).⁸³ AI may be helpful in those instances, and there is value in further research being conducted in this area. From a legal standpoint, applying England's laws as an example, the conclusion that can be drawn at present is that, where a dispute or question arises for a

⁷⁷Desai, *Violence and surveillance in mental health wards*, *supra* note 53.

⁷⁸Pia Puolakka & Steven Van De Steene, *Artificial Intelligence in Prisons in 2030. An exploration on the future of Artificial Intelligence in Prisons*, 11 ADVANCING CORRECTIONS JOURNAL 128, 131 (2021), <https://rm.coe.int/ai-in-prisons-2030-acjournal/1680a40b83> [<https://perma.cc/4W3X-GP4R>]; James Redden et al., *Artificial Intelligence Applications in Corrections*, U.S. DEP'T JUST. at 1, 4 (2020), <https://cjtec.org/files/5f5f9458ebc72> [<https://perma.cc/7UKL-TUX3>].

⁷⁹Greer et al., *supra* note 75, at 2.

⁸⁰Puolakka & Steene, *supra* note 78; similar systems are used in the UK, see Cara McGoogan, *Liverpool prison using AI to stop drugs and weapons smuggling*, THE TELEGRAPH, Dec. 6, 2016, <https://www.telegraph.co.uk/technology/2016/12/06/liverpool-prison-using-ai-stop-drugs-weapons-smuggling/> [<https://perma.cc/4ABG-BAS4>].

⁸¹Puolakka & Steene, *supra* note 78.

⁸²Chris Francescani, *US prisons and jails using AI to mass-monitor millions of inmate calls*, ABCNEWS (Oct. 24, 2019), <https://abcnews.go.com/Technology/us-prisons-jails-ai-mass-monitor-millions-inmate/story?id=66370244> [<https://perma.cc/34N6-VWHP>].

⁸³Due et al., *supra* note 72.

court to consider, a surveillance system in mental health wards may require approval from a court and may be subject to constant review by a judge. However, from a practical standpoint, those legal protections may not always manifest, meaning privacy concerns will persist. In most cases, it appears that surveillance is used without questions arising for a court to consider, and there is a risk that AI-enabled systems will fall within the same void despite the greater legal questions or risks those systems may pose owing to their pervasive nature.

3.2 Predictive Behavior Models for Suicide

Many patients are detained under mental health laws for treatment, and many of them pose a considerable risk to themselves or others.⁸⁴ Given the frequency of violence and suicide attempts in psychiatric units, accurate prediction of such behaviors is critical for proper treatment.⁸⁵ The current structure for treatment relies on predictive models that are rarely specific to an individual, instead “classifying a patient to a certain group and referring to that group’s averages.”⁸⁶ Traditional statistical models accommodate a limited number of variables, limiting predictions based on surrogate markers of suicide attempts, suicidal planning and ideation, and self-injury.⁸⁷ Psychiatric units also use risk assessment instruments to predict aggressive behavior.⁸⁸ These instruments use individual medical and social history data, treatment conditions, behaviors, and psychopathology.⁸⁹ Yet, these tools are “based on one-off assessments of risk factors, use electronic or paper forms, and only predict risk over the very short-term (i.e. the next 24 hours).”⁹⁰ Moreover, these factors become clinically irrelevant in an emergency where the medical history of a possibly violent detained patient is unknown.⁹¹ Other common methods include analyzing EHRs, national registries and self-reported questionnaires, although the latter’s reliability is questionable due to the patient’s motivation, capacity, and self-awareness.⁹²

Overall, the predictive ability of suicide has not improved in fifty years of research.⁹³ Suicide prediction remains incorrect in most cases.⁹⁴ Yet eighty-three percent of people who die by suicide have contact with health services the year before their death, and forty-five percent have contact the month before.⁹⁵ Therefore, there is a significant opportunity to assist doctors in assessing the risk of suicide using AI when those patients present.⁹⁶

⁸⁴Alvaro Barrera et al., *Introducing artificial intelligence in acute psychiatric inpatient care: qualitative study of its use to conduct nursing observations*, 23 *BMJ MENTAL HEALTH* 34, 34 (2020).

⁸⁵Arne E Vaaler et al., *Short-term prediction of threatening and violent behaviour in an Acute Psychiatric Intensive Care Unit based on patient and environment characteristics*, 11 *BMC PSYCH.* 1 (2011).

⁸⁶Paulina Cecula et al., *Applications of artificial intelligence to improve patient flow on mental health inpatient units – Narrative literature review* *HELIYON*, Apr. 2021, at 1, 3, <https://doi.org/10.1016/j.heliyon.2021.e06626> [<https://perma.cc/76R3-58U8>].

⁸⁷Charlotte Wells et al., *Artificial Intelligence and Machine Learning in Mental Health Services: An Environmental Scan*, *CADTH HEALTH TECH. REV.* at 1, 23 (2021), https://www.cadth.ca/sites/default/files/attachments/2021-06/artificial_intelligence_and_machine_learning_in_mental_health_services_environmental_scan.pdf.

⁸⁸Taanvi Ramesh et al., *Use of Risk Assessment Instruments to Predict Violence in Forensic Psychiatric Hospitals: a Systematic Review and Meta-Analysis*, 52 *EUR. PSYCH.* 47, 47 (2018).

⁸⁹Vaaler et al., *supra* note 85, at 1.

⁹⁰Seena Fazel et al., *Modifiable Risk Factors for Inpatient Violence in Psychiatric Hospital: Prospective Study and Prediction Model*, *PSYCHOL. MED.* 590, 590 (2021).

⁹¹Vaaler et al., *supra* note 85, at 2.

⁹²Cecula et al., *supra* note 86, at 6.

⁹³Daniel D’Hotman & Erwin Loh, *AI enabled suicide prediction tools: a qualitative narrative review*, 27 *BMJ HEALTH & CARE INFORMATICS*, Oct. 2020, at 1, 2 (citing Franklin JC et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. 143 *PSYCHOL. BULL.* 187 (2017)).

⁹⁴Alan L. Berman & Gregory Carter, *Technological Advances and the Future of Suicide Prevention: Ethical, Legal, and Empirical Challenges*, 50 *SUICIDE AND LIFE-THREATENING BEHAV.* 643, 644 (2019).

⁹⁵D’Hotman & Loh, *supra* note 92, at 2 (citing Ahmedani BK, Simon GE, Stewart C, et al. Health care contacts in the year before suicide death, 29 *J. GEN. INTERN. MED.* 870 (2014).

⁹⁶*Id.* at 2.

AI could be used in hospitals as a decision support for assessing suicide risk using surveillance or by examining EHRs, hospital records, and other government data sources.⁹⁷ Inpatients may also provide data using smartphones to log their mood, sleep patterns, step count, and interactions with other patients.⁹⁸ At the same time, using such systems could pose risks for patients in two realms: first, in research, when AI systems are being developed; second, in the ward, once AI has been deployed.

On research in the mental health context generally (not only inpatients), Marks has examined AI-based suicide prediction algorithms and the risks they pose.⁹⁹ Marks highlights AI tools that examine medical records to predict suicide and social interactions that fall outside the health care sphere.¹⁰⁰ Focusing on the former, academic medical centres, hospitals, and government agencies usually examine medical records containing information about speciality and primary care visits.¹⁰¹ In the United States, suicide prediction is governed by the Health Insurance Portability and Accountability Act (HIPAA) which protects privacy and imposes penalties on entities when patient data is breached.¹⁰² Research that uses patient data for suicide prediction is also subject to the Federal Common Rule, medical ethics rules and IRBs, all designed to protect human research subjects.¹⁰³ These avenues have their own challenges. AI is autonomous, so dynamic consent may be needed from patients to use their data.¹⁰⁴ Consent is also imperative where public data is integrated into predictive models.¹⁰⁵

Additional considerations will arise where AI is approved for use for predicting suicide on a mental health ward in the future. Marks highlights specific risks that arise with such technology and groups those risks into concerns about safety, privacy, and autonomy.¹⁰⁶ That framework is helpful in the inpatient context. Safety risks include false negatives and false positives.¹⁰⁷ Those misclassifications could identify someone as being at risk of suicide (when they are not), or not at risk of suicide (when they are).¹⁰⁸ The resulting action will be wrong for that patient. A patient thought to not be at risk may attempt suicide. A patient thought to be at risk may be detained and lose their freedoms when they are not at risk.¹⁰⁹ Those patients may also risk having suicidal thoughts as part of their permanent medical record when it is not true. Moreover, involuntary admission in these circumstances could be traumatic and dehumanizing, raising the risk of suicide.¹¹⁰ Improper detention or the denial of access to clinical services could form the basis of a legal claim against the clinician, the health care entity, or the developer.¹¹¹

A potential factor in this paradigm is the complexity in mental health disorders, which have highly complex aetiology and symptomology that do not lend well to strict classification, thus, affecting the reliability of predictive algorithms.¹¹² In practice, small sample sizes for psychiatric disorders are used

⁹⁷*Id.* at 1.

⁹⁸See Alina Haines-Delmont *et al.*, *Testing Suicide Risk Prediction Algorithms Using Phone Measurements With Patients in Acute Mental Health Settings: Feasibility Study*, 8 JMIR MHEALTH UHEALTH 1 (2020), <https://mhealth.jmir.org/2020/6/e15901/PDF>; see also Ashley Jane Bruen *et al.*, *Exploring Suicidal Ideation Using an Innovative Mobile App-Strength Within Me: The Usability and Acceptability of Setting up a Trial Involving Mobile Technology and Mental Health Service Users*, 7 JMIR MENTAL HEALTH, Feb. 2020, 1, <https://mental.jmir.org/2020/9/e18407/PDF>.

⁹⁹Mason Marks, *Artificial Intelligence-Based Suicide Prediction*, 18 YALE J. L. & TECH. 98, 102 (2019).

¹⁰⁰*Id.* at 105.

¹⁰¹*Id.* at 106.

¹⁰²*Id.* at 105.

¹⁰³*Id.* at 105.

¹⁰⁴Berman & Carter, *supra* note 94, at 645.

¹⁰⁵See Raymond Tucker *et al.*, *Ethical and practical considerations in the use of a predictive model to trigger suicide prevention interventions in healthcare settings*, 49 SUICIDE AND LIFE-THREATENING BEHAV. 382, 387 (2019).

¹⁰⁶Marks, *supra* note 99, at 111.

¹⁰⁷*Id.*

¹⁰⁸*Id.* at 111-12; see also, Jordan M. Braciszewski, *Digital Technology for Suicide Prevention*, 1 ADVANCES IN PSYCH. AND BEHAV. HEALTH 53, 55 (2021).

¹⁰⁹Marks, *supra* note 99, at 112.

¹¹⁰*Id.* at 113.

¹¹¹Berman & Carter, *supra* note 94, at 644; see also Price *et al.*, *supra* note 30.

¹¹²Wells *et al.*, *supra* note 87, at 27-28.

which limits data for training and testing, leading to overfitting of the data and skewed accuracy measurements. Clinicians are forced to rely on less accurate models, adversely affecting suicide prediction.

While doctors should exercise their judgement, they may be reluctant to ignore the AI system because of concerns about medical liability.¹¹³ It is widely held in psychiatric practice that suicide assessment tools should be used mainly to aid clinical judgment based on a clinical, interview-based risk assessment. However, the fear of legal consequences could result in clinicians relying more on AI than clinical judgement, which could threaten clinicians losing psychiatric risk assessment skills. Practice will vary in hospitals and jurisdictions, but the fear of liability may lead some doctors to take extra measures based on poorly assessed risks—measures that are costly in terms of the personnel required. An inpatient deemed high risk for suicide could be placed under constant direct observation, with frequent assessments being completed. That task may be cumbersome and unnecessary, but clinicians may nevertheless “play it safe” without worrying about resource implications. Thus, in a paradigm where AI exists in psychiatric wards, there are misclassification risks that could lead to staff taking inappropriate action for that patient. AI might improve risk prediction when trained properly but could provide poorer predictive capabilities than traditional methods owing to poor training or input data coupled with an overreliance and trust placed in those systems by doctors.

Privacy concerns cover data breaches and the transfer or sale of data to third parties, such as advertisers and data brokers, leading to affected individuals being stigmatized, exploited or discriminated against.¹¹⁴ In the United States, HIPAA protects patient privacy by requiring that suicide-related data cannot be transmitted without being deidentified.¹¹⁵ Health care providers are also prohibited from sharing non-anonymized health data with advertisers.¹¹⁶ In England, strict rules on surveillance were noted covering the length of time that data can be stored, who can access and the restriction on third parties using that data. For AI in a psychiatric ward, strict data rules could significantly restrict many systems from being used or restrict how they handle data if they are used. If AI is used on wards, it is unclear how these factors will manifest on the ground. At present, it is hard to safeguard the confidentiality of data used in digital mental health tools due to apps being poorly understandable and developers having opaque data management policies.¹¹⁷ Such a lack of clarity could not be permitted in the mental health setting.

Autonomy covers risks of censorship, unnecessary confinement or civil commitment, and criminal penalties or incarceration in countries where suicide attempts are illegal.¹¹⁸ Warrantless searches could be conducted on individuals at risk of suicide based on the AI system’s prediction.¹¹⁹ In a psychiatric ward, a patient could be moved into a locked ward based on a misclassification, undermining their autonomy, be traumatic and inadvertently increase the risk of harm. This implicates consent because confinement could occur without consent based on erroneous information. In reality, the patient is not at risk, and it would be an egregious breach of their rights if confined.¹²⁰

Overall, AI may help to overcome some existing limitations of predicting suicide. However, its use is not without risk. Doctors may be reluctant to disagree with AI recommendations, but that reluctance could lead to inappropriate decisions for the patient, resulting in the loss of autonomy (where the AI system is wrong). Whether AI could be used in practice owing to privacy considerations restricting the use of patient data is unclear.

¹¹³Marks, *supra* note 99, at 111–112.

¹¹⁴*Id.* at 111; see also Braciszewski, *supra* note 108; see also John Torous et al., *Smartphones, Sensors, and Machine Learning to Advance Real-Time Prediction and Interventions for Suicide Prevention: a Review of Current Progress and Next Steps*, 20 CURR. PSYCH. REP. Jan.–Dec. 2018, at 4, <https://doi.org/10.1007/s11920-018-0914-y> [<https://perma.cc/FJ9E-RB26>].

¹¹⁵Marks, *supra* note 99, at 117.

¹¹⁶*Id.*

¹¹⁷See, Nicole Martinez-Martin et al., *Ethics of Digital Mental Health During COVID-19: Crisis and Opportunities*, 7 JMIR MENTAL HEALTH Dec. 2020, at 4, <https://mental.jmir.org/2020/12/e23776/PDF>.

¹¹⁸Marks, *supra* note 99, at 120.

¹¹⁹*Id.*

¹²⁰On informed consent risks, see Braciszewski, *supra* note 108, at 59.

3.3 AI and Hospital Administration

In some hospitals, staff can be overburdened with administrative work, reducing the time for direct patient care. Bed and staff shortages often result in suboptimal risk assessments and early discharges, with subsequent high readmission rates. This part highlights these added challenges of inefficient bed management and mismatch of staff competencies to patient needs due to scheduling irregularities. It further explores the use of AI in ward administration and the legal risks associated with its incorporation in mental health.

While these administrative challenges may exist in other wards, one cannot ignore their impact in psychiatry where there is “severe underinvestment” in many parts of the world.¹²¹ Bed shortages in mental health wards have increased over the years, and worsened during COVID-19, causing a supply-demand mismatch.¹²² Inadequate bed management causes difficulties in procuring beds when needed, resulting in frequent patient transfers that disorient patients suffering from certain mental disorders, leading to stress and loss of trust in services.¹²³ Moreover, the risk of homelessness and incarceration of patients with severe mental illnesses and increased risks of physical harm to other patients or members of the public, are all possible consequences of insufficient psychiatric bed capacity.¹²⁴

Traditionally, hospital managers oversee capacity and demand based on estimations and historical knowledge. Currently, bed registries and bed tracking systems are used to track bed availability in psychiatric units. However, since the registries are not linked to EHRs or hospital admission/discharge data systems, bed tracking systems are not updated automatically. Instead, they rely on administrative and medical personnel to update manually.¹²⁵ Time-consuming meetings are often held to discuss case management, provide updates on predicted discharges, and the current demand for beds. Information is not always up-to-date due to the staff's focus on other responsibilities. The registries also struggle to match bed availability to patients with complex needs, such as those who exhibit violence or aggression, have co-occurring medical conditions or needs.¹²⁶ This leads to suboptimal risk assessment, with the likelihood of patients being discharged before they receive proper treatment, resulting in increased readmission rates.¹²⁷

Similarly, staff shortages and a mismatch of staff competencies to patient needs in psychiatric wards pose a significant barrier to treating mental health disorders due to improper utilization of staff expertise. Understaffing and poor staff rota design has led to unclear and overlapping job responsibilities, which adds to the inefficiencies in psychiatric units.¹²⁸ According to the BMA, “staff working in mental health care are at ‘breaking point’ as they try to handle rising demand with a continuous staffing rota gap.”¹²⁹ The copious administrative work allows little time for one-on-one direct patient care time. Long working days (of up to twelve hours) and inadequate staff scheduling have been identified as a significant threat to

¹²¹Faraaz Mahomed, *Addressing the Problem of Severe Underinvestment in Mental Health and Well-Being from a Human Rights Perspective*, 22 HEALTH HUM. RTS J. 35 (2020), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7348439/pdf/hhr-22-01-035.pdf>.

¹²²Adrian James, *Mental health beds are full, leaving patients without treatment and clinicians with difficult choices*, BMJ OP. (Mar. 5, 2021), <https://blogs.bmj.com/bmj/2021/03/05/mental-health-beds-are-full-leaving-patients-without-treatment-and-clinicians-with-difficult-choices/> [<https://perma.cc/M4P2-DCC8>].

¹²³Tejas Kotwal et al., *Bed Management in Psychiatry: Ensuring That the Patient Perspective Is Not Forgotten*, 27 BJPSYCH ADVANCES 352, 352 (2021).

¹²⁴Eric P. Slade & Howard H. Goldman, *The Dynamics of Psychiatric Bed Use in General Hospitals*, 42 ADMIN. POL'Y MENTAL HEALTH 139, 139 (2015).

¹²⁵Tami L. Mark et al., *Bed Tracking Systems: Do They Help Address Challenges in Finding Available Inpatient Beds?*, 70 PSYCH. SERV. 921, 921-922 (2019).

¹²⁶*Id.* at 924.

¹²⁷Dawoodbhoy, *supra* note 3, at 14.

¹²⁸*Id.*

¹²⁹Elisabeth Mahase, *Workforce crisis has left mental health staff at “breaking point” as demand rises*, THE BMJ (Jan. 9, 2020), <https://doi.org/10.1136/bmj.m88> [<https://perma.cc/A9TV-79PC>].

employee health and well-being, affecting their mental health.¹³⁰ These rota gaps adversely impact staff training, morale, and the quality of care they provide. Consequently, medical personnel are often “forced to act above their competencies, putting patient safety at risk.”¹³¹

While bed shortages and staff rota challenges exist in other wards, the problems in psychiatric wards are at “crisis” levels.¹³² Hospitals, in general, are shifting towards using AI to assist with these administrative challenges. For bed management, AI is being deployed to predict discharges to free up bed space. A Portugal-based company developed an AI solution that predicts discharges by combining data about bed usage with clinical data to suggest actions five to seven days ahead. The AI system considers ward gender, surgery schedules, patient age and condition, type of bed needed, number of beds and nursing schedules, among other variables, increasing bed manager productivity by thirty to fifty percent.¹³³ This reduces stress and enables staff to identify and prioritize the sickest patients, make faster decisions, and allocate hospital resources efficiently.

Similarly, AI is being used to predict practitioners’ workload in ICUs and EDs and ensure that the right skilled workers are linked with appropriate shifts.¹³⁴ These systems extract information, match the competencies of health care workers to specific tasks, and help fill available slots, making the planning of rotas more efficient. AI systems are also designed to consider policy requirements when making recommendations, such as law-mandated working hours or the presence of senior doctors for specific shifts.¹³⁵

These solutions can be implemented in psychiatric wards, albeit with caution. Developing a successful AI system for psychiatric ward administration requires high-quality data to train the algorithm and validate the model.¹³⁶ Given that this sensitive medical data is a regulated asset, risks arise over data accuracy and privacy.¹³⁷ A case study involving AI used to improve hospital bed allocation at Kettering General Hospital revealed the challenges of obtaining enough good quality data that cover the complexity of patient needs. These challenges include the need for the technology to be reconfigured regularly with new information about increased beds, changed ward layouts or flu admission peaks.¹³⁸

In mental health care, the challenges of obtaining enough data and the use of EHRs to train resource allocation and staff scheduling algorithms are amplified and somewhat circular. AI requiring data on psychiatric inpatients must comply with data protections noted above. For example, in the United States, health care data in EHRs is considered sensitive, with mental health data being even more sensitive due to the high risk of stigmatization and discrimination.¹³⁹ The HIPAA Privacy Rule provides special protection for mental health information, including extra protection for psychotherapy notes and

¹³⁰Joanna Pryce et al., *Evaluation of an open-rota system in a Danish psychiatric hospital: A mechanism for improving job satisfaction and work-life balance*, 14 J. NURSING MGMT. 282, 283 (2006).

¹³¹BRITISH MED. ASS’N, MEDICAL ROTA GAPS IN ENGLAND, 2 (2018), <https://www.bma.org.uk/media/3550/medical-rota-gaps-in-england-2018.pdf>.

¹³²See AM. PSYCH. ASS’N, THE PSYCHIATRIC BED CRISIS IN THE US: UNDERSTANDING THE PROBLEM AND MOVING TOWARD SOLUTIONS (May 2022), <https://www.psychiatry.org/getmedia/81f685f1-036e-4311-8dfc-e13ac425380f/APA-Psychiatric-Bed-Crisis-Report-Full.pdf>.

¹³³Nuno Forneas, *Improving hospital bed management with AI*, IBM (Aug. 16 2018), <https://www.ibm.com/blogs/client-voices/improving-hospital-bed-management-ai/> [https://perma.cc/ZE9G-WZMJ].

¹³⁴Liu, *supra* note 17.

¹³⁵How Hospitals Can Tap AI To Manage Staff Better Amid Covid-19 Crisis, *supra* note 18.

¹³⁶WORLD HEALTH ORGANIZATION [WHO], ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH, 36 (2021).

¹³⁷Matthias Klumpp et al., *Artificial Intelligence for Hospital Health Care: Application Cases and Answers to Challenges in European Hospitals*, 9 HEALTHCARE, July 2021, at 15-16, 20, <https://doi.org/10.3390/healthcare9080961> [https://perma.cc/W8DP-QR4Y].

¹³⁸Improving hospital bed allocation using AI, NHS (Dec. 10, 2021), <https://transform.england.nhs.uk/ai-lab/explore-all-resources/develop-ai/improving-hospital-bed-allocation-using-ai/> [https://perma.cc/QZ9X-EY6D].

¹³⁹Christopher A. Lovejoy, *Technology and mental health: The role of artificial intelligence*, 55 EUR. PSYCH. Jan. 2019 at 2, <https://doi.org/10.1016/j.eurpsy.2018.08.004> [https://perma.cc/4WC5-ZW5X].

personal notes contained within, and requires the patient's written authorization for its use or disclosure.¹⁴⁰

However, even if research is legally compliant, access to data may be difficult owing to the lack of data, the extra-sensitive nature of the information, and the restrictions on who it can be shared with. The quality of the data on the EHR may also be insufficient for adequately training the algorithm. Mental health information is "regularly missing from EHRs, documented in the wrong place, or under documented in specific contexts."¹⁴¹ This is likely due to the subjective nature of information,¹⁴² greater reliance on narrative progress notes as opposed to quantitative measures,¹⁴³ and stigma surrounding mental health conditions.¹⁴⁴ Fragmented use of paper and electronic records also contributes to missing mental health information on EHRs.¹⁴⁵ This ambiguous and, at times, incorrect documentation of mental health data may be insufficient to train AI systems for the administration of psychiatric wards. Thus, even if developers can overcome the legal challenge surrounding privacy, they may be limited by the data that is available.

4. The Need for Guidelines

The analysis above reveals that although AI-based solutions have shown promise for monitoring patients, predicting illness, and assisting with hospital administration, the application of the technology in psychiatric wards may be challenging. The risks associated with adopting such technology require serious scrutiny, given the capacity of AI recommendations to aggravate existing issues or to provide limited assistance. The legal and ethical implications highlighted surrounding patient privacy, informed consent, access to accurate data, and data protection and privacy might curtail the adoption of AI systems in the psychiatric setting. This part further elaborates on these considerations and calls for guidelines to be developed that account for those risks should AI systems eventually be used in psychiatric wards.¹⁴⁶ Some AI systems may never gain approval for use, but the purpose here is to provide pre-emptive baseline considerations for stakeholders instead of reactionary measures post-implementation.

First, a major concern is the loss of privacy for patients. For example, delusions of persecution, in the form of camera monitoring, are among the most common delusions in mental health services. Many patients perceive being subjected to constant monitoring as dehumanizing, and continuous surveillance increases stress and anxiety levels in patients and decreases trust in those around them,¹⁴⁷ which may worsen their existing mental illness and make them more resistant towards adopting surveillance technology for their care.¹⁴⁸ Guidelines for AI video surveillance should be adopted to protect privacy by mandating that patient data should be used for a specified purpose and shared with only a narrow range of people, such as their doctor on the ward or the developers, to ensure the AI system operates

¹⁴⁰U.S. DEP'T HEALTH AND HUMAN SERV., HIPAA PRIVACY RULE AND SHARING INFORMATION RELATED TO MENTAL HEALTH, 2, <https://www.hhs.gov/sites/default/files/hipaa-privacy-rule-and-sharing-info-related-to-mental-health.pdf> (last visited Jan 5, 2022).

¹⁴¹Timothy Charles Kariotis et al., *Impact of Electronic Health Records on Information Practices in Mental Health Contexts: Scoping Review*, 24 J. MED. INT. RES. 10 (2022).

¹⁴²*Id.* at 10.

¹⁴³*Id.* at 23.

¹⁴⁴*Id.* at 10.

¹⁴⁵*Id.*

¹⁴⁶Guidelines are often used as a soft law 'stop gap' for gaps in AI governance until hard laws are developed. For a framework presenting this process, see Barry Solaiman, *From AI to Law in Healthcare: The Proliferation of Global Guidelines in a Void of Legal Uncertainty*, 42 MED. AND LAW 391 (2023).

¹⁴⁷A.J. Marsden & William Nesbitt, *I Spy With My Little Eye: The origins and effects of mass surveillance*, PSYCHOLOGY TODAY (Nov. 6, 2017), <https://www.psychologytoday.com/us/blog/myth-the-mind/201711/i-spy-my-little-eye> [<https://perma.cc/JK4G-XQW7>].

¹⁴⁸Nasriah Zakaria & Rusyaizila Ramli, *Physical factors that influence patients' privacy perception toward a psychiatric behavioral monitoring system: a qualitative study*, 14 NEUROPSYCHIATRIC DISEASE AND TREATMENT 117, 118 (2018).

properly. Smart cameras can be configured to employ privacy mask applications to hide or blur out faces in videos or anything else that might reveal a person's health information, such as computer screens displaying patient details.¹⁴⁹ This would allow enhanced monitoring of psychiatric patients in their rooms while lowering the risk of violating their privacy and ensuring that their sensitive health information remains secure. From a research perspective, any research must comply with heightened norms and protections surrounding mental health data. However, data is crucial to training effective algorithms. One potential solution could be to develop algorithms that can be trained on anonymous pixels instead of patient images.¹⁵⁰

Second, underlying the privacy concerns is the need for consent for using AI systems in psychiatric wards. While patients' consent to being monitored by AI in a public ward is already a contested issue, it is further complicated in a psychiatric ward where patients are more likely to be admitted involuntarily.¹⁵¹ The stigma associated with mental illnesses and its potential effect on a patient's professional and personal life – work opportunities or marriage prospects, for instance – marks a significant challenge for obtaining consent for using AI systems in that context. Furthermore, even if consent were obtained for AI, that consent might be rendered void where an AI system evolves by making recommendations beyond its initial intended purpose that the patient agreed to.

Adequately informing the patients and hospital staff when using an AI system and the type of data these systems have access to is not enough. Approval may be required from the courts for using or affirming the use of AI systems on the ward and be subject to ongoing review by a judge. A corollary challenge is explaining the reasons for AI recommendations (explainability and the black-box problem) to doctors and patients.¹⁵² As such, future guidelines should require that the relevant legal approvals be obtained where necessary for the use of an AI system and that those systems should be explainable so that stakeholders can understand the capabilities, scope of use, and limitations of these systems.¹⁵³

Further, should consent challenges be overcome, it may be necessary to implement a flexible approach for obtaining consent in practice, owing to the ever-evolving nature of AI systems. One option could be electronic informed consent to improve individual control and choice.¹⁵⁴ It would allow users to understand the consequences of consent by determining which data will be used and how. Given the frequent updates of AI software, this approach can be enhanced by incorporating "dynamic consent," allowing patients (with capacity) to modify their consent periodically depending on the use of the information they wish to permit.¹⁵⁵ A framework that allows patients to opt out of constant surveillance could also be considered. Taking Eindhoven-based Catharine Hospital, for instance, the hospital and researchers obtained explicit consent and designed the smart camera with a switch to turn off the device if a patient did not wish to be monitored.

Third, considerations around data security, privacy and algorithmic accuracy arise. For example, health data collected by AI may exceed what is required and may later be repurposed for uses that have serious ethical, legal, and human rights implications.¹⁵⁶ Massive amounts of mental health data are required to train AI algorithms for suicide prediction and for the effective administration of psychiatric wards. Policies

¹⁴⁹James Thorpe, *Exclusive: Using AI to improve healthcare operations*, INT'L SEC. J. (Jan. 22, 2021), <https://internationalsecurityjournal.com/using-ai-to-improve-healthcare/> [<https://perma.cc/ZN2X-WR64>].

¹⁵⁰*Smart camera reduces unexpected complications in hospitals*, *supra* note 11.

¹⁵¹Appenzeller et al., *supra* note 51, at 483.

¹⁵²Barry Solaiman & Mark Bloom, *AI, Explainability and Safeguarding Patient Safety in Europe: Toward a Science-Focused Regulatory Model* in *THE FUTURE OF MEDICAL DEVICE REGULATION: INNOVATION AND PROTECTION* 91 (I Glenn Cohen et al. eds. 2022).

¹⁵³David D. Luxton et al., *Ethical Issues and Artificial Intelligence Technologies in Behavioral and Mental Health Care*, in, *ARTIFICIAL INTELLIGENCE IN BEHAVIORAL AND MENTAL HEALTH CARE* 264 (David D. Luxton ed., 2016).

¹⁵⁴For the distinction between informed consent and obtaining consent in AI, see, Barry Solaiman, *Addressing Access with Artificial Intelligence: Overcoming the Limitations of Deep Learning to Broaden Remote Care Today*, 51 THE U. OF MEMPHIS L. REV. 1103, 1131-38, (2020).

¹⁵⁵WHO, *supra* note 136, at 82.

¹⁵⁶*Id.* at 43.

regulating such data collection should address mechanisms aimed at preventing the “collection of illicit or unverified sensitive data.”¹⁵⁷ Given this sensitivity of data, clear guidelines are needed highlighting the laws that must be followed when handling such data in these and other scenarios.¹⁵⁸ Clear frameworks for how medical practitioners and researchers use data while safeguarding patient confidentiality will promote public trust, advancing the use of AI in the health care setting.

5. Conclusion

Overall, the potential use of AI in psychiatric wards raises challenges that are distinct enough within the mental health domain to warrant closer consideration. The examples of AI integration into surveillance systems, suicide prediction, and hospital administration highlight those specificities concerning patient privacy, informed consent, and data protection, among others. While those concerns may also arise in other hospital wards, additional factors apply in the mental health care arena owing to the greater sensitivity of the information there. For that reason, laws, norms, practices, and protections have developed to protect psychiatric patients – factors that ultimately trickle down into considerations about using AI in psychiatric wards.

The examination in this Article yields two pertinent insights. First, should AI be approved for use in systems on psychiatric wards, any such approval would likely be contingent upon significant restrictions regarding how those systems will be used and monitored. Further, if those systems are used, they will likely be on an individual patient basis and be subject to frequent review where a legal question arises. Second, specific guidance on the legal considerations that arise would be beneficial for hospitals, patients, and the courts to establish best practices for using and implementing AI on mental health wards. Without such guidance, risks may arise of different hospitals using and implementing such systems incongruently, resulting in diverging levels of protection for patients. That ought to be the next contemplation for research in this area, supported by interdisciplinary efforts between hospitals, developers, lawyers, and bioethicists.

Barry Solaiman is the Assistant Dean of Students and Assistant Professor of Law at Hamad Bin Khalifa University (HBKU), Qatar, and an Adjunct Assistant Professor of Ethics in Clinical Medicine at Weill Cornell Medicine – Qatar. He holds a PhD in Law from the University of Cambridge and a Fellowship in Bioethics from Harvard Medical School’s Center for Bioethics. Solaiman is co-editor of a major book titled ‘Health, AI and the Law’, and has published widely in the field. He leads major research projects on regulating AI in health. He was formerly Editor-in-Chief of both the Cambridge International Law Journal and the Medicine and Law Journal.

Abeer Malik is a Research Governance & Compliance Specialist at the HBKU Office of Vice President for Research in Qatar. She holds a Juris Doctor degree from HBKU Law, honored with the Dean’s Award for academic excellence. Prior to that, she received the “Order of Excellence” for her Bachelors in Business Administration from Qatar University. Her research interests span across healthcare law, the intersection of law and technology, and international human rights law.

Suhaila Ghuloum is a Senior Consultant Psychiatrist at Hamad Medical Corporation, Professor at Doha Institute for Graduate Studies, and an Associate Professor of clinical psychiatry at Weill Cornell Medicine – Qatar. She has national and international key involvements in mental health, including with the GCC, WHO, the Arab Board, the Royal College of Psychiatrists, the World Psychiatric Association, and the World Federation for Mental Health. Dr. Ghuloum has several research projects and publications.

¹⁵⁷Filippo Pesapane et al., *Legal and Regulatory Framework for AI Solutions in Healthcare in EU, US, China, and Russia: New Scenarios after a Pandemic*, 1 RADIATION 261, 270 (2021).

¹⁵⁸WHO, *supra* note 136, at 90.