
CONTINUING EDUCATION

Ethics and Governance of Artificial Intelligence in Health Care: A Continuing Education Manual for Psychologists

Reading Manual for Licensed Psychologists
and Graduate-Level Clinicians

FRANKLIN CARVAJAL

August 2026

Contents

Abstract	5
Introduction	6
Historical and Theoretical Foundations	7
The Ethical Lineage of Health Technology	7
Foundations in Bioethics and Human Rights	8
The Role of Regulation and Governance	8
Conceptual Models and Mechanisms	9
Defining Artificial Intelligence in the Health Context	9
The Concept of Biomedical Big Data	9
Categories of AI Applications in Health	10
Mechanisms of Benefit and Harm	10
Integrated Review of the Evidence	11
Convergent Themes Across the Guidance	11
Points of Tension and Unresolved Questions	12
Synthesis Table: Mapping WHO Principles to Psychological Practice	12
Professional Interpretation and Critical Analysis	13
Strengths of the WHO Framework	13
Limitations and Critical Considerations	14
The Meaning of the Evidence for Psychology	15
Psychological Pathways and Stress Responses	15
The Psychological Impact of AI on Patients	15
The Psychological Impact of AI on Clinicians	17
Mental Health Outcomes and Severity Spectrum	18
AI Applications Across the Continuum of Mental Health Care	18
Severity and the Question of Clinical Adequacy	19

System-Level and Contextual Psychological Effects	20
AI, Health Systems, and the Distribution of Care	20
The Digital Divide and Global Inequities	20
Surveillance, Autonomy, and the Public Good	21
Clinical Implications for Psychologists	22
Rethinking Assessment in an AI-Augmented Environment	22
Formulation and Treatment Planning	23
Intervention Delivery	24
Outcome Monitoring and Quality Improvement	24
Ethical Considerations in Practice	25
Competence in an Era of AI	25
Informed Consent	25
Confidentiality and Data Protection	26
Multiple Relationships and Commercial Interests	26
Cultural Responsiveness	27
Scope of Practice	27
Practice Recommendations	28
Well-Supported Recommendations	28
Tentative Recommendations	29
Future Directions and Research Gaps	31
The Evolving Landscape	31
Empirical Evidence on AI in Mental Health	31
The Development of Practice Guidelines	31
Governance and Regulation	32
Training and Workforce Development	32
Cross-Disciplinary Collaboration	32
Attention to Global Equity	32

Unresolved Ethical Tensions	33
Conclusion	33
References	34

Abstract

Artificial intelligence (AI) is rapidly reshaping the landscape of health care, and psychology is not exempt from this transformation. From algorithmic risk stratification in psychiatric settings, to chatbot-based mental health interventions, to machine-learning tools that assist in diagnostic formulation, psychologists are increasingly encountering AI in ways that touch the core of their professional practice. This continuing education manual draws upon the World Health Organization's guidance on the ethics and governance of AI for health (World Health Organization [WHO], 2021) to provide licensed psychologists and graduate-level clinicians with a scholarly, in-depth examination of the ethical, clinical, and systemic dimensions of AI in health-care delivery. The manual explores six foundational ethical principles articulated by the WHO Expert Group on Ethics and Governance of AI for Health: protecting human autonomy; promoting human well-being, safety, and the public interest; ensuring transparency, explainability, and intelligibility; fostering responsibility and accountability; ensuring inclusiveness and equity; and promoting AI that is responsive and sustainable (WHO, 2021). Each principle is examined through the lens of psychological practice, with attention to how it maps onto the *Ethical Principles of Psychologists and Code of Conduct* and to the practical realities of contemporary clinical work. Illustrative case scenarios are integrated throughout to demonstrate how these principles operate in real-world decision-making. The manual concludes with concrete practice recommendations, an analysis of research gaps, and a discussion of how psychologists can position themselves as ethically informed stewards of AI-augmented care.

Introduction

For much of its history, the profession of psychology has grounded itself in a fundamentally human enterprise: the sustained, disciplined attention of one person to the inner life of another. Assessment, formulation, therapeutic alliance, and outcome monitoring have all relied on the clinician's capacity to observe, interpret, and respond to the unique presentation of each individual. The introduction of artificial intelligence into health care disturbs this picture in ways that are both promising and unsettling. Algorithms now screen electronic health records for suicide risk, natural language processing tools parse therapy transcripts, wearable devices infer affective states from physiological signals, and large language models offer conversational support that many patients experience as therapeutic. Whether these developments enhance or erode the quality of psychological care depends, in large part, on the ethical scaffolding that surrounds them.

The World Health Organization released its comprehensive guidance document, *Ethics and Governance of Artificial Intelligence for Health*, in 2021, following extensive deliberation by a 20-member Expert Group drawn from public health, medicine, law, human rights, technology, and ethics. The guidance identifies AI as holding "great promise for the practice of public health and medicine" while emphasizing that realizing this promise depends on collective effort to design and implement ethically defensible laws, policies, and technologies (WHO, 2021). The document is not aimed exclusively at physicians or public health officials; its scope explicitly includes anyone who designs, finances, regulates, or uses AI technologies for health. Psychologists occupy several of these roles simultaneously. They use AI-informed clinical tools, they interpret AI-generated outputs, they may consult on the design of behavioral health technologies, and they are increasingly asked to evaluate the psychological consequences of AI deployment in schools, workplaces, and health systems.

This manual is written for licensed psychologists and graduate-level clinicians who wish to develop a sophisticated, ethically grounded understanding of AI in health care. It does not assume prior technical training in computer science, though it does presume familiarity with core clinical concepts, ethical reasoning, and the structure of contemporary health systems. The manual synthesizes the WHO guidance into a coherent narrative, extends its principles into the specific domain of psychological practice, and offers practical recommendations that clinicians can apply in their daily work. Throughout, the manual takes seriously both the potential benefits of AI and the very real risks it poses to patients, communities, and the integrity of the

psychological professions.

The stakes of getting this right are considerable. AI systems, once embedded in health-care workflows, are difficult to remove. They shape the categories through which clinicians perceive patients; they influence what is documented and what is overlooked; they alter the distribution of professional attention and resources. If these systems encode biases, they will propagate those biases at scale. If they are opaque, they will erode the informed consent that lies at the heart of psychological practice. If they concentrate power in the hands of a few technology companies, they will reshape the political economy of mental health care in ways that may not serve patients or communities well. Conversely, if psychologists engage thoughtfully with AI, contribute to its design, and advocate for its ethical governance, the profession has an opportunity to help ensure that these powerful tools genuinely advance human flourishing.

Historical and Theoretical Foundations

The Ethical Lineage of Health Technology

The ethical questions raised by AI in health care are, in an important sense, not new. As the WHO guidance notes, "many of the ethical concerns described in this report predate the advent of AI, although AI itself presents a number of novel concerns" (WHO, 2021). The introduction of any powerful new technology into medicine has historically forced a reexamination of core ethical commitments. The development of psychopharmacology in the mid-twentieth century raised questions about the boundary between treatment and social control. The rise of managed care in the 1980s and 1990s prompted debates about the compatibility of commercial incentives with clinical fidelity. The digitization of medical records in the 2000s reopened questions about privacy, confidentiality, and the ownership of health information. AI extends and intensifies these debates but does not create them from nothing.

What is distinctive about AI is the combination of scale, opacity, and adaptivity. Scale, because a single algorithm can influence millions of clinical decisions across many jurisdictions. Opacity, because the internal workings of many machine-learning systems, particularly deep neural networks, are not readily interpretable even by their designers. Adaptivity, because AI systems that continue to learn from new data can shift their behavior over time in ways that may not be transparent to users. These features distinguish AI from earlier medical technologies and require ethical frameworks that go beyond traditional bioethics.

Foundations in Bioethics and Human Rights

The WHO guidance situates its ethical analysis within two intersecting traditions: bioethics and human rights. Bioethics, as it has developed since the mid-twentieth century, has produced a set of widely recognized principles—respect for autonomy, beneficence, non-maleficence, and justice—that continue to inform medical and psychological ethics. Human rights frameworks, articulated most prominently in the Universal Declaration of Human Rights and subsequent international instruments, provide legally and morally binding commitments to dignity, equality, privacy, and freedom from discrimination. The WHO guidance draws on both traditions, arguing that AI for health must be evaluated not only against clinical standards of safety and efficacy but also against broader commitments to human rights and social justice (WHO, 2021).

For psychologists, these traditions are already deeply familiar. The American Psychological Association's ethical code emphasizes beneficence and non-maleficence, fidelity and responsibility, integrity, justice, and respect for people's rights and dignity. These principles map closely onto the WHO framework, though the WHO document extends them in ways that are particularly relevant to the digital era, adding explicit attention to transparency, explainability, and the sustainability of health systems.

The Role of Regulation and Governance

A third foundation is regulatory and governance theory. The WHO guidance emphasizes that ethical principles alone are insufficient; they must be operationalized through laws, policies, institutional practices, and mechanisms of accountability. "Governance in health," as the guidance defines it, refers to "a range of functions for steering and rule-making by governments and other decision-makers, including international health agencies, to achieve national health policy objectives conducive to universal health coverage" (WHO, 2021). This governance perspective is essential for understanding why AI ethics cannot be reduced to individual clinical decisions. The choices made by regulators, ministries of health, hospital administrators, and technology companies shape the environment in which individual clinicians act. Psychologists who wish to practice ethically in an AI-augmented world must attend to these system-level factors as well as to the immediate clinical encounter.

Conceptual Models and Mechanisms

Defining Artificial Intelligence in the Health Context

The WHO guidance offers a deliberately non-technical definition of AI as "the ability of algorithms encoded in technology to learn from data so that they can perform automated tasks without every step in the process having to be programmed explicitly by a human" (WHO, 2021). This definition captures the essential feature that distinguishes contemporary AI from earlier forms of computerized decision support: the capacity for learning from data rather than following explicit rules written by programmers. Machine learning, a subset of AI, encompasses a family of techniques including supervised learning (in which algorithms learn to map inputs to labeled outputs), unsupervised learning (in which algorithms identify patterns in unlabeled data), and reinforcement learning (in which algorithms learn through trial and error to maximize a reward signal).

For psychologists, the practical implication of this definition is that AI systems are not neutral tools that simply execute human intentions. They embody choices—about what data to use, what outcomes to optimize, what patterns to detect—that were made by their designers and that shape what the system will and will not see. An AI system trained to predict psychiatric readmission using data from one health system may perform poorly, and even harmfully, when applied to a population whose demographic composition, clinical presentations, or documentation practices differ from those of the training data.

The Concept of Biomedical Big Data

Closely related to AI is the concept of "big data," which the WHO guidance treats as a foundational resource for machine-learning applications in health. Biomedical or health big data encompasses electronic health records, medical imaging archives, genomic sequences, insurance claims data, wearable device outputs, social media content, and administrative records from health systems (WHO, 2021). The aggregation of these heterogeneous data streams enables the training of AI models that would have been impossible a generation ago. It also creates new ethical vulnerabilities. Data that were collected for one purpose—clinical documentation, insurance billing, consumer marketing—may be repurposed for AI training in ways that the original data subjects never anticipated or consented to.

For psychology, the implications are particularly acute. Mental health data are among the most sensitive categories of personal information. Their inclusion in large

training datasets, especially without robust consent and de-identification protocols, raises significant concerns about privacy, stigma, and the potential for discrimination.

Categories of AI Applications in Health

The WHO guidance provides a non-comprehensive classification of AI applications in health, including uses in clinical medicine, health research, drug development, health-systems management and planning, and public health surveillance (WHO, 2021). Within each of these domains, AI is being deployed in ways that intersect with psychological practice. In clinical medicine, AI supports diagnostic decision-making, treatment planning, and outcome prediction. In health research, AI accelerates the identification of patterns in large datasets that might inform new theories of psychopathology or intervention. In health-systems management, AI is used to allocate resources, schedule appointments, and forecast demand for services. In public health surveillance, AI monitors populations for emerging health threats, including mental health crises.

Each of these applications carries distinct ethical implications. A diagnostic AI that operates within an individual clinical encounter raises different concerns than a surveillance AI that operates at the population level. Understanding these differences is essential for psychologists who wish to evaluate specific AI tools they may encounter in their practice.

Mechanisms of Benefit and Harm

The WHO guidance identifies several mechanisms by which AI can benefit health care. AI can augment clinician capacity by handling routine cognitive tasks, freeing time for higher-value activities. It can improve diagnostic accuracy by detecting patterns that exceed human perceptual capacities. It can optimize treatment plans by integrating large numbers of variables. It can support pandemic preparedness by modeling disease spread. It can extend care to underserved populations by lowering the marginal cost of certain services (WHO, 2021).

The same guidance is equally clear about the mechanisms of harm. AI can encode and amplify biases present in training data. It can undermine clinician judgment by fostering over-reliance on algorithmic recommendations. It can erode privacy through the aggregation and analysis of personal data. It can concentrate power in the hands of technology developers who face weak accountability. It can enable surveillance and social control in ways that violate human rights. It can displace

human workers, including health-care workers, in ways that have both economic and psychological consequences. And it can consume substantial energy resources, contributing to environmental harm (WHO, 2021).

These mechanisms of benefit and harm are not independent; they often operate simultaneously in the same application. An AI-based mental health chatbot may expand access to psychological support for people who cannot afford or access traditional therapy, while simultaneously raising concerns about privacy, quality of care, and the substitution of algorithmic support for human relationship. The task of ethical evaluation is not to declare AI good or bad in the abstract but to assess specific applications in specific contexts.

Integrated Review of the Evidence

Convergent Themes Across the Guidance

The WHO guidance, though a single document, integrates the perspectives of a diverse expert group and reflects a broad consensus about the state of the field. Several convergent themes emerge. First, there is broad agreement that AI holds substantial promise for improving health outcomes, particularly in settings where human resources are constrained. Second, there is equally broad agreement that this promise is contingent on ethical design and deployment, and that ethical failures can produce serious harms. Third, there is convergence on the view that ethical principles alone are insufficient; they must be operationalized through governance structures that include laws, regulations, professional standards, and mechanisms for public participation.

The six ethical principles articulated in the guidance—protecting human autonomy; promoting human well-being, safety, and the public interest; ensuring transparency, explainability, and intelligibility; fostering responsibility and accountability; ensuring inclusiveness and equity; and promoting AI that is responsive and sustainable—represent the core substantive commitments of this consensus (WHO, 2021). Each principle addresses a distinct ethical concern, but the principles are interrelated. Autonomy cannot be protected without transparency; equity cannot be ensured without accountability; sustainability requires both responsiveness and inclusiveness.

Points of Tension and Unresolved Questions

The guidance also identifies points of tension that remain unresolved. One such tension concerns the relationship between innovation and precaution. Rapid deployment of AI can bring benefits sooner but also increases the risk of undetected harms. Slower, more cautious deployment reduces risk but may delay benefits, particularly for underserved populations. Another tension concerns the balance between global standards and local adaptation. AI systems developed in high-income countries may not perform well in low- and middle-income settings, but developing separate systems for each context is resource-intensive and may perpetuate technological dependence. A third tension concerns the role of commercial actors. Private technology companies possess the expertise and capital to develop sophisticated AI systems, but their commercial interests may not align with public health priorities.

These tensions do not admit of simple resolution. The WHO guidance treats them as ongoing challenges to be navigated through deliberative governance rather than as problems to be solved once and for all. This posture is important for psychologists to understand. Ethical practice in an AI-augmented world is not a matter of applying a fixed set of rules; it requires ongoing judgment, dialogue, and adaptation.

Synthesis Table: Mapping WHO Principles to Psychological Practice

The table below offers an original synthesis mapping the six WHO principles onto specific domains of psychological practice.

WHO Ethical Principle	Core Concern	Application to Psychological Practice
Protecting human autonomy	Preserving human control over decisions	Ensuring informed consent for AI-augmented assessment; maintaining clinician judgment over algorithmic recommendations
Promoting well-being, safety, and public interest	Preventing harm and ensuring benefit	Verifying that AI tools meet safety and efficacy standards for the specific patient population served

WHO Ethical Principle	Core Concern	Application to Psychological Practice
Ensuring transparency, explainability, intelligibility	Making AI understandable to stakeholders	Explaining to patients how AI contributes to their assessment or treatment; understanding the basis of algorithmic recommendations
Fostering responsibility and accountability	Clarifying who is answerable when things go wrong	Understanding the psychologist's professional responsibility for AI-assisted decisions; establishing mechanisms for redress
Ensuring inclusiveness and equity	Preventing bias and expanding access	Evaluating whether AI tools perform equitably across demographic groups; advocating for access in underserved settings
Promoting responsiveness and sustainability	Ongoing evaluation and long-term viability	Monitoring AI performance in one's own practice; considering environmental and workforce impacts

Note. Synthesized from WHO (2021).

Professional Interpretation and Critical Analysis

Strengths of the WHO Framework

The WHO guidance represents a significant contribution to the international discourse on AI ethics in health care. Its strengths are substantial. It offers a principled framework that is grounded in both bioethics and human rights, giving it broad ethical legitimacy. It is developed through a deliberative process involving diverse experts, lending it procedural legitimacy. It addresses both individual clinical encounters and system-level governance, providing conceptual scope. And it is oriented toward action, with practical recommendations directed at specific stakeholder groups (WHO, 2021).

For psychologists, several features of the guidance are particularly valuable. The emphasis on human autonomy resonates deeply with the profession's longstanding commitment to respect for persons. The attention to transparency and explainability aligns with the psychologist's obligation to provide informed consent and to make

clinical reasoning accessible to patients. The focus on inclusiveness and equity reinforces the profession's growing attention to cultural responsiveness and health disparities. And the recognition that AI raises system-level issues extends the psychologist's ethical horizon beyond the consulting room.

Limitations and Critical Considerations

The guidance also has limitations that psychologists should recognize. It is a high-level document that identifies principles and general directions rather than providing detailed technical or clinical protocols. It cannot anticipate every specific AI application, and its principles must be interpreted and applied in particular contexts. It reflects a consensus among experts, which necessarily involves compromises and may underweight minority perspectives, particularly those of communities most affected by AI deployment. And it addresses AI in health broadly, without specific attention to the distinctive features of mental health care.

Several critical observations are worth making. First, the guidance's emphasis on human oversight and human warranty, while appropriate, may be difficult to operationalize as AI systems become more complex and more embedded in clinical workflows. The cognitive burden on clinicians to critically evaluate algorithmic recommendations can be substantial, and the risk of automation bias—the tendency to over-rely on automated systems—is well documented in the human factors literature. Simply asserting that humans should remain "in the loop" does not resolve the practical challenge of making that oversight meaningful.

Second, the guidance's treatment of commercial interests is somewhat abstract. It acknowledges that patients "require assurance that their rights and interests will not be subordinated to the powerful commercial interests of technology companies" (WHO, 2021), but the mechanisms for providing such assurance remain underdeveloped. In practice, the commercial development of AI in health care is dominated by a small number of large technology companies whose accountability to patients and clinicians is limited.

Third, the guidance's discussion of the digital divide, while important, may understate the depth of the challenge. The gap between well-resourced health systems that can afford sophisticated AI infrastructure and under-resourced systems that cannot is not merely a matter of technology transfer; it reflects broader structural inequalities that are unlikely to be resolved by AI-specific interventions.

Fourth, the guidance does not directly engage with the specific ethical challenges of AI in mental health care, where issues of stigma, self-identity, and the therapeutic relationship take particular forms. Psychologists must therefore adapt and extend the general principles to their specific practice domain.

The Meaning of the Evidence for Psychology

Taken together, the WHO guidance suggests that psychology as a profession stands at an important juncture. The tools of AI will increasingly touch every aspect of psychological practice, from screening and assessment to intervention and outcome monitoring. Psychologists can respond to this development in several ways. They can adopt AI tools uncritically, deferring to technology developers and institutional pressures. They can resist AI wholesale, protecting the traditional forms of practice at the cost of foreclosing potential benefits. Or they can engage thoughtfully, developing the technical literacy and ethical judgment needed to evaluate specific tools in specific contexts, and contributing their professional expertise to the design and governance of AI in health care.

The third path is the one that the WHO guidance implicitly commends and that this manual seeks to support. It requires psychologists to develop new competencies while maintaining their traditional ones. It requires the profession collectively to engage with policy, regulation, and industry in ways that may be unfamiliar. And it requires an honest reckoning with both the potential benefits of AI and the very real risks it poses to patients, communities, and the profession itself.

Psychological Pathways and Stress Responses

The Psychological Impact of AI on Patients

Although the WHO guidance focuses primarily on ethical principles and governance frameworks, its analysis has important implications for understanding the psychological pathways through which AI affects patients. When patients encounter AI in their care—whether through a chatbot, an algorithmic risk assessment, or a machine-learning-based diagnostic tool—they experience specific psychological responses that psychologists are uniquely positioned to understand and address.

One important pathway involves trust and perceived agency. The WHO guidance emphasizes that patients "require assurance that their rights and interests will not be subordinated" to commercial or governmental interests (WHO, 2021). When patients feel that decisions about their care are being made by opaque algorithms

rather than by clinicians who know them, they may experience a loss of agency that can undermine engagement with treatment. This effect is particularly pronounced in mental health settings, where the sense of being genuinely heard and understood is often central to therapeutic benefit.

A second pathway involves privacy concerns and the stress associated with data exposure. Patients who learn that their psychological information may be used to train AI systems, or that their behavior is being monitored by algorithmic tools, may experience heightened vigilance and reduced disclosure. This can degrade the quality of assessment and intervention, creating a feedback loop in which the presence of AI paradoxically reduces the quality of the data on which it relies.

A third pathway involves the internalization of algorithmic categorizations. When patients are told that an AI system has identified them as being at high risk for a particular outcome, this categorization can become part of their self-concept in ways that may be helpful or harmful. Labeling effects have long been studied in psychology, and AI adds new dimensions to this phenomenon by producing categorizations that carry the weight of computational authority while often lacking transparent grounding.

Case Scenario 1: Initial Presentation and Algorithmic Risk

Dr. M is a licensed psychologist practicing in a large integrated health system. She receives a referral for a 34-year-old patient, Mr. J, who was flagged by the system's electronic-health-record-based suicide risk prediction algorithm as being in the highest risk category. The algorithm's output includes a numerical risk score but does not indicate which features drove the prediction. Mr. J was informed at intake that "our system uses tools to help identify patients who might benefit from additional support," but he was not told the specifics of the algorithmic assessment.

When Dr. M meets with Mr. J, he presents with moderate depressive symptoms and stress related to a recent job change, but no active suicidal ideation and no clear risk factors that Dr. M can identify from the interview. Mr. J asks Dr. M why he was referred to her specifically. Dr. M faces several ethical decisions: whether and how to disclose the role of the algorithm in generating the referral, how to weigh the algorithmic risk score against her own clinical assessment, and how to document her formulation.

Analysis. This scenario illustrates the tension between the WHO principles of transparency and clinical judgment. The principle of transparency, as articulated in the guidance, requires that patients understand the role AI systems play in their care (WHO, 2021). Dr. M's obligations under the Ethical Principles of Psychologists and Code of Conduct similarly emphasize informed consent and honesty. At the same time, Dr. M must exercise clinical judgment about whether the algorithmic categorization is accurate and how it should influence her treatment planning. If she defers uncritically to the algorithm, she risks over-treating a patient whose risk may be overestimated. If she dismisses it, she risks under-attending to a patient whose risk may exceed what her interview reveals. The path forward involves transparent communication with Mr. J about the referral pathway, careful clinical assessment that neither privileges nor dismisses the algorithmic input, and documentation that reflects her integrated formulation.

The Psychological Impact of AI on Clinicians

The WHO guidance also implicitly addresses the psychological consequences of AI for health-care workers, though this receives less sustained attention. When clinicians encounter AI in their workflows, they may experience effects that shape their professional functioning. These include cognitive load associated with evaluating algorithmic recommendations, moral distress when institutional pressures push toward reliance on tools the clinician judges inadequate, and identity threat when AI is framed as potentially replacing clinical judgment.

The guidance's attention to workforce impacts and to the need for training reflects an awareness of these dynamics. It notes that governments and companies must "address anticipated disruptions in the workplace, including training for health-care workers to adapt to the use of AI systems, and potential job losses due to use of automated systems" (WHO, 2021). For psychologists, these considerations extend to questions about how AI will reshape the profession itself—what tasks will be automated, what new competencies will be required, and how the meaning of psychological practice will be redefined.

Mental Health Outcomes and Severity Spectrum

AI Applications Across the Continuum of Mental Health Care

Applications of AI span the full continuum of mental health care, from preventive screening in general populations to intensive treatment for severe psychiatric conditions. At the preventive end, AI-based screening tools deployed in primary care, schools, and workplaces aim to identify individuals at risk for common mental health conditions before they reach clinical severity. At the treatment end, AI supports diagnostic formulation, treatment selection, and outcome prediction for patients with established psychiatric diagnoses. Beyond the traditional treatment system, AI-based consumer applications offer symptom tracking, self-help interventions, and conversational support directly to users without professional involvement.

Each of these applications raises distinct ethical considerations. Preventive screening at scale can identify individuals who might benefit from earlier intervention, but it can also generate false positives that cause unnecessary distress and consume clinical resources. The WHO guidance emphasizes that AI applications should be evaluated for "safety, accuracy and efficacy for well-defined use cases or indications" (WHO, 2021), a standard that is often more difficult to meet for screening applications than for tools that support clinical decisions about already-identified patients.

Diagnostic and treatment-planning AI raises questions about the appropriate weight to give algorithmic recommendations relative to clinical judgment. The WHO guidance's emphasis on human warranty—the requirement that AI development and deployment include meaningful evaluation by clinicians and patients—is particularly relevant here. In mental health care, where diagnostic categories are contested and where treatment response is difficult to predict, uncritical reliance on algorithmic recommendations carries substantial risks.

Consumer-facing AI applications, including mental health chatbots and self-guided digital interventions, present the most complex ethical picture. They may expand access to support for people who cannot access traditional care, but they operate largely outside the regulatory frameworks that govern professional practice. The WHO guidance's emphasis on responsibility and accountability is challenging to apply when the "provider" of care is a consumer application whose ownership, data practices, and clinical validation may be opaque.

Case Scenario 2: The Chatbot and the Adolescent

Dr. R is a psychologist who specializes in adolescent mental health. During an intake with a 16-year-old patient, S, Dr. R learns that S has been using a popular AI chatbot for "emotional support" for approximately six months. S reports finding the chatbot helpful because "it's always there" and "doesn't judge me." S also acknowledges that during a period of acute distress two months earlier, they told the chatbot about passive suicidal thoughts, and the chatbot's response was empathic but did not clearly connect them to crisis resources. S did not tell their parents or any other person about these thoughts at the time.

Dr. R must consider how to integrate this information into the clinical picture, whether to discuss the limitations of chatbot support with S, whether and how to involve the family, and how to think about S's ongoing use of the chatbot alongside her professional treatment.

Analysis. This scenario illustrates several dimensions of the WHO framework simultaneously. The principle of promoting well-being and safety is engaged by the question of whether the chatbot's response to S's suicidal ideation was safe and appropriate. The principle of accountability is engaged by the question of who, if anyone, is responsible for the adequacy of that response. The principle of autonomy is engaged by S's genuine preference for chatbot support and by their right to make choices about their own care. Dr. R's clinical task is not to prohibit chatbot use—which would likely be both ineffective and paternalistic—but to help S develop a thoughtful understanding of what such tools can and cannot do, to strengthen S's connection to human support systems, and to model transparent conversation about the role of AI in mental health.

Severity and the Question of Clinical Adequacy

A recurring question across the severity spectrum is whether AI tools are adequate to the clinical situation in which they are deployed. The WHO guidance frames this in terms of "well-defined use cases or indications" (WHO, 2021), a standard that requires clarity about what a tool is designed to do, for whom, and under what conditions. For low-severity applications, such as psychoeducation or symptom tracking, this standard may be relatively straightforward to meet. For high-severity applications, such as crisis response or treatment of severe psychiatric conditions, the standard is far more demanding.

Psychologists play an important role in evaluating this adequacy. Their clinical training equips them to assess whether a specific tool is fit for a specific purpose in a specific context. This evaluative role is one of the profession's distinctive contributions to the ethical integration of AI into mental health care.

System-Level and Contextual Psychological Effects

AI, Health Systems, and the Distribution of Care

Beyond individual clinical encounters, AI is reshaping the systems within which psychological care is delivered. The WHO guidance emphasizes that AI can "inform the decisions of health policy-makers or allocate resources within health systems" (WHO, 2021). In mental health care, algorithmic tools are increasingly used to triage referrals, prioritize waiting lists, allocate limited treatment slots, and determine service intensity. These system-level applications can have profound psychological consequences even when individual clinicians never see the algorithm's output directly.

Consider a health system that uses an algorithm to determine which patients on a mental health waiting list should be seen most urgently. The algorithm's design—what features it uses, what outcomes it optimizes, how it handles missing data—shapes who gets seen first. Patients whose characteristics are well-represented in the training data may be prioritized appropriately, while patients from underrepresented groups may be systematically deprioritized. The psychological consequences of delayed care, including symptom worsening and reduced engagement, fall disproportionately on the deprioritized groups.

The Digital Divide and Global Inequities

The WHO guidance devotes substantial attention to the digital divide, warning that AI systems "based primarily on data of individuals in high-income countries may not perform well for individuals in low- and middle-income settings" (WHO, 2021). This concern extends beyond geographic boundaries. Within any given country, AI systems trained on data from well-resourced patient populations may perform poorly for marginalized groups, including racial and ethnic minorities, rural populations, and those with limited English proficiency.

For mental health care specifically, the digital divide takes several forms. Access to AI-based interventions is unevenly distributed, with well-resourced patients able to purchase premium consumer applications while under-resourced patients rely on lower-quality free options. The performance of diagnostic and predictive AI varies across demographic groups, with underrepresented populations often experiencing higher error rates. And the cultural assumptions embedded in AI-based interventions may not align with the needs and preferences of diverse communities.

Surveillance, Autonomy, and the Public Good

The WHO guidance raises important concerns about the use of AI for surveillance and social control. In mental health contexts, these concerns take particular forms. AI systems that monitor social media for suicide risk, workplaces for signs of psychological distress, or school populations for indications of behavioral problems can identify individuals in need of support, but they also create infrastructures of monitoring that may chill authentic expression and self-disclosure. The guidance's emphasis on the potential for "unwarranted surveillance" as a misuse of AI (WHO, 2021) is directly applicable.

The COVID-19 pandemic, discussed briefly in the guidance, offered a preview of these dynamics. Proximity tracking applications, symptom-monitoring tools, and predictive analytics all raised questions about the appropriate boundaries between public health surveillance and individual privacy (WHO, 2021). Analogous questions arise in mental health contexts, particularly as AI tools become embedded in schools, workplaces, and community settings.

Case Scenario 3: Workplace Wellness and Algorithmic Monitoring

Dr. L is a psychologist providing consulting services to a large employer that is considering adopting an AI-based "workplace wellness" platform. The platform monitors employee communications, calendar patterns, and self-reported mood ratings to identify employees who may be experiencing psychological distress. The platform's vendor claims that this monitoring will enable earlier intervention and reduce absenteeism and turnover. The employer wants Dr. L's professional opinion on whether to adopt the platform.

Dr. L must consider multiple issues: the accuracy and validity of the platform's algorithms, the implications for employee privacy and autonomy, the risk that identification of distress could lead to discriminatory treatment, the adequacy of the follow-up services offered to identified employees, and the broader effect of algorithmic monitoring on workplace culture and psychological safety.

Analysis. This scenario engages nearly every principle in the WHO framework. The principle of autonomy is engaged by the question of whether employees can meaningfully consent to monitoring in an employment context where refusal may carry professional consequences. The principle of transparency is engaged by the need to understand what the platform actually monitors and how its algorithms function. The principle of equity is engaged by the risk that algorithmic identification of distress could disadvantage employees from groups already marginalized in the workplace. The principle of responsibility is engaged by the question of who is accountable if the platform's outputs are misused. Dr. L's role as a consultant is not simply to endorse or reject the platform but to help the employer think through these issues systematically, to identify what conditions would need to be met for ethical adoption, and, where appropriate, to advocate for alternatives that better serve employee well-being.

Clinical Implications for Psychologists

Rethinking Assessment in an AI-Augmented Environment

The integration of AI into health care changes what psychological assessment means and how it should be conducted. Traditional psychological assessment relies on the clinician's structured observation and integration of information from interview, testing, and collateral sources. AI adds new sources of information—algorithmic risk scores, natural-language-processing outputs, physiological data from wearables—and new potential biases and errors. Psychologists conducting assessment in AI-augmented environments must develop the capacity to interpret these new information sources critically, to understand their limitations, and to integrate them with traditional assessment methods.

Several practical implications follow. First, psychologists should routinely inquire about their patients' use of AI-based tools, including consumer mental health applications, chatbots, and self-monitoring devices. This information is often not volunteered but can be clinically significant. Second, psychologists should be attentive to algorithmic outputs that appear in referrals, medical records, and consultation notes, treating them as one source of information among others rather than as authoritative determinations. Third, psychologists should document their own reasoning clearly, including how they weighed algorithmic inputs relative to other information sources.

Formulation and Treatment Planning

Case formulation—the integrative narrative that connects a patient's presentation to underlying mechanisms and to intervention strategy—is a distinctively human contribution that psychologists bring to care. AI can support formulation by identifying patterns in data, but it does not substitute for the interpretive work of formulation itself. The WHO guidance's emphasis on human control over medical decisions (WHO, 2021) supports the view that formulation should remain a fundamentally clinician-led process, informed but not determined by algorithmic tools.

In treatment planning, similar considerations apply. AI-based recommendation systems may suggest particular interventions based on patterns in outcome data, but treatment selection involves considerations—patient preferences, therapeutic alliance, cultural fit, contextual constraints—that algorithms typically cannot integrate. Psychologists should treat AI recommendations as hypotheses to be evaluated rather than as prescriptions to be followed.

Case Scenario 4: Diagnostic Discordance

Dr. T is conducting an initial evaluation of a 42-year-old woman referred by her primary care physician. The referral notes that a diagnostic support tool used in the primary care practice suggested a diagnosis of bipolar II disorder based on the patient's self-reported symptoms. During the evaluation, Dr. T finds that the patient's symptom pattern is more consistent with major depressive disorder with anxious features and a history of significant trauma. The patient reports that she was told by her primary care physician that "the computer thinks you have bipolar."

Dr. T must decide how to communicate her diagnostic impression to the patient, how to address the previous diagnostic communication, and how to coordinate with the referring physician.

Analysis. This scenario highlights the risks of algorithmic diagnostic tools when their outputs are communicated to patients as authoritative determinations. The WHO principles of transparency and explainability are engaged by the question of what the patient was told and understood about the basis of the earlier diagnostic communication. The principle of responsibility is engaged by the question of clinical accountability for that communication. Dr. T's task is to conduct her assessment on its own terms, to communicate her impression to the patient with appropriate humility about diagnostic uncertainty, and to coordinate professionally with the referring physician in a way that supports better use of diagnostic tools in the future.

Intervention Delivery

AI is increasingly integrated into the delivery of psychological interventions themselves. Digital cognitive-behavioral therapy programs, AI-augmented therapy platforms, and chatbot-based interventions represent varying degrees of algorithmic involvement in the therapeutic process. The WHO guidance's emphasis on evaluating tools against "well-defined use cases or indications" (WHO, 2021) is particularly important here. An AI-based intervention that is effective for mild-to-moderate depression in adults may be inappropriate or harmful for other populations or conditions.

Psychologists who incorporate AI-based interventions into their practice should apply the same standards of evidence-based care that they would apply to any other intervention. This includes reviewing the empirical evidence for the specific tool, considering its fit for the specific patient, monitoring outcomes carefully, and being prepared to discontinue use if the tool is not producing the expected benefit.

Outcome Monitoring and Quality Improvement

The WHO guidance emphasizes the importance of ongoing evaluation of AI systems in actual use (WHO, 2021). This principle of responsiveness has direct implications for how psychologists incorporate AI into their practice. Rather than adopting AI tools and assuming they will function as advertised, psychologists should establish local practices for monitoring outcomes, identifying problems, and adjusting use accordingly. This may include tracking whether AI-informed decisions are associated with better or worse outcomes for the psychologist's patient population, attending to disparities in tool performance across demographic groups, and being alert to any patterns of harm.

Ethical Considerations in Practice

Competence in an Era of AI

The ethical principle of competence, foundational to psychological practice, takes on new dimensions in an AI-augmented environment. Psychologists have long been expected to practice within the boundaries of their training and experience. The introduction of AI tools raises questions about what competence in the use of such tools requires. It is unlikely that most psychologists need to become experts in machine learning or in the technical details of specific algorithms. It is likely, however, that psychologists need to develop what might be called critical AI literacy: the capacity to understand what AI tools do and do not do, to evaluate claims about their performance, to communicate with patients about their role in care, and to recognize when their use is or is not appropriate.

This form of competence is not static. As AI tools evolve, psychologists' understanding must evolve with them. Continuing education, professional consultation, and engagement with the emerging scholarly literature are all essential. The WHO guidance's emphasis on training for health-care workers is directly relevant (WHO, 2021), though the responsibility for developing this training falls on multiple actors: employers, professional associations, and individual clinicians themselves.

Informed Consent

Informed consent is a cornerstone of psychological practice, and AI complicates it in several ways. Patients must be able to understand what they are consenting to, but AI systems are often opaque even to their developers. Consent must be voluntary, but patients may face situations in which declining to consent to AI-augmented care effectively means declining care altogether. And consent must be specific, but the uses of data in AI systems may be difficult to specify in advance.

The WHO guidance's emphasis on the "protection of privacy and confidentiality and obtaining valid informed consent through appropriate legal frameworks for data protection" (WHO, 2021) points to the importance of consent processes that go beyond formalities. Meaningful consent in an AI-augmented environment requires that patients understand, in accessible terms, what AI is being used in their care, what data are being collected and how they may be used, what alternatives are available, and what the risks and potential benefits are. Developing consent processes that meet this standard is an ongoing challenge for the profession.

Confidentiality and Data Protection

Confidentiality is central to psychological practice, and the data-intensive nature of AI creates new challenges for maintaining it. When patient information is used to train AI systems, when clinical records are analyzed by algorithmic tools, or when patient interactions with AI-based interventions are logged and stored, the boundaries of confidentiality become more complex. Psychologists must understand where patient data are going, who has access to them, and what safeguards are in place.

The WHO guidance emphasizes that data protection requires appropriate legal frameworks (WHO, 2021), but psychologists cannot rely solely on legal frameworks. They must also exercise professional judgment about what data practices are compatible with the trust relationship at the heart of clinical work. In some cases, this may mean declining to use tools whose data practices are inadequate; in other cases, it may mean advocating for stronger protections within institutional and professional contexts.

Multiple Relationships and Commercial Interests

The WHO guidance raises concerns about the commercial pressures that can distort the development and deployment of AI in health care (WHO, 2021). For psychologists, these concerns intersect with the ethical prohibition on harmful multiple relationships and with the broader obligation to avoid conflicts of interest. Psychologists who receive compensation from AI vendors, who develop AI tools themselves, or who work in settings where AI adoption is driven by commercial rather than clinical considerations, must attend carefully to how these relationships shape their professional judgment.

Transparency with patients about such relationships is essential. So is a willingness to prioritize patient welfare over institutional or commercial pressures. The WHO guidance's caution that "companies and governments should introduce AI technologies only to improve the human condition and not for objectives such as unwarranted surveillance or to increase the sale of unrelated commercial goods and services" (WHO, 2021) sets a high standard that individual psychologists can support through their own choices and advocacy.

Cultural Responsiveness

The WHO guidance's emphasis on inclusiveness and equity, and its warning about AI systems that may not perform well outside high-income settings (WHO, 2021), aligns with the psychology profession's growing attention to cultural responsiveness. Psychologists who use AI tools should routinely consider whether those tools have been validated for the populations they serve, whether they perform equitably across demographic groups, and whether their use may inadvertently disadvantage patients from marginalized communities.

This concern extends beyond individual tool selection to broader questions about the design of AI in mental health. Psychologists who consult on AI development have opportunities to advocate for training data that better represents diverse populations, for validation studies that examine performance across groups, and for design choices that reflect the diversity of human experience.

Scope of Practice

Finally, the introduction of AI raises questions about the psychologist's scope of practice. What counts as psychological practice when much of the routine work of assessment, monitoring, or intervention delivery is handled by algorithms? What is the psychologist's responsibility for the outputs of tools they use but did not develop? These questions do not have simple answers, but they must be engaged. The WHO guidance's emphasis on accountability (WHO, 2021) suggests that psychologists cannot disclaim responsibility for AI-informed decisions on the grounds that "the algorithm did it." Professional responsibility remains with the human clinician who uses the tool, and psychologists must be prepared to exercise and defend that responsibility.

Case Scenario 5: The Ambitious Employer

Dr. K works as an in-house psychologist for a health system that has recently invested heavily in AI-based mental health screening. The system's leadership has directed that all patients seen in primary care should be screened using an AI tool that generates automated risk profiles, and that psychologists should prioritize their limited appointment slots based on these profiles. Dr. K has concerns about the accuracy of the tool for the system's diverse patient population and about the ethical implications of algorithmic prioritization, but she is aware that pushing back may create professional difficulties.

Analysis. This scenario illustrates how individual ethical practice is constrained by institutional context. Dr. K's obligations include respect for patient autonomy, provision of competent care, and avoidance of practices that may harm patients or perpetuate inequities. The WHO guidance's emphasis on accountability (WHO, 2021) suggests that Dr. K cannot simply defer to institutional decisions if she believes they will harm patients. At the same time, unilateral refusal to use the tool may not be effective and may compromise her ability to advocate for change from within. A thoughtful response might involve documenting her concerns clearly, seeking evidence about the tool's performance in her patient population, engaging colleagues and leadership in dialogue about the concerns, exercising clinical judgment about how to integrate the tool's outputs with her own assessment, and advocating through professional channels for stronger evaluation and governance of AI in the system.

Practice Recommendations

The following recommendations are offered as concrete guidance for psychologists navigating the integration of AI into their practice. They are grouped by domain and distinguished by level of support: well-supported recommendations are grounded directly in the WHO guidance and in established ethical principles; tentative recommendations extend these principles into areas where evidence and consensus are still developing.

Well-Supported Recommendations

Develop critical AI literacy. Every psychologist should invest in developing a working understanding of what AI is, how it functions in health care, and what its limitations are. This does not require technical expertise in machine learning, but it does require the capacity to evaluate claims about AI tools critically. The WHO guidance's emphasis on the need for education and training for health-care workers (WHO, 2021) supports this recommendation.

Inquire about patients' use of AI-based tools. Routinely ask patients about their use of consumer mental health applications, chatbots, self-monitoring devices, and other AI-based tools. This information is often clinically significant and is unlikely to be volunteered without inquiry.

Practice transparent communication about AI in care. When AI tools contribute to a patient's assessment or treatment, communicate this clearly and in accessible terms. Support the patient's understanding of what the tool does, what its limitations

are, and how it relates to the clinician's own judgment. This recommendation follows directly from the WHO principle of transparency (WHO, 2021).

Maintain clinical judgment as central. Treat AI outputs as one source of information among others, not as authoritative determinations. Integrate algorithmic inputs with interview data, testing results, collateral information, and clinical formulation. Document your reasoning clearly, including how you weighed different sources of information.

Attend to data protection. Understand where patient data go when AI tools are used, who has access to those data, and what safeguards are in place. Decline to use tools whose data practices are inadequate, and advocate for stronger protections within institutional contexts.

Evaluate tools for the populations you serve. Before adopting an AI tool, consider whether it has been validated for the population you will use it with. Be alert to potential biases that may disadvantage patients from marginalized groups. This aligns directly with the WHO principle of equity (WHO, 2021).

Document professional responsibility. When AI tools contribute to clinical decisions, document your own reasoning and the basis for your judgment. Do not attribute clinical decisions solely to algorithmic outputs. Maintain the professional accountability that patient care requires.

Tentative Recommendations

Contribute to AI governance where possible. Psychologists have distinctive expertise that can inform the ethical governance of AI in health care. Where opportunities arise—through professional associations, institutional committees, regulatory processes, or research—consider contributing that expertise. This recommendation extends the WHO framework's emphasis on multi-stakeholder governance (WHO, 2021) into a specific professional call.

Develop practice-level monitoring of AI outcomes. As AI tools become embedded in practice, develop local practices for monitoring their effects. Track whether AI-informed decisions are associated with expected outcomes; be alert to patterns of harm or inequity. This recommendation reflects the WHO principle of responsiveness (WHO, 2021) but requires further development of specific methods for implementation.

Advocate for professional standards. As AI tools proliferate in mental health care, advocate for professional standards that govern their use. Support the development of guidelines by professional associations, and engage in dialogue with colleagues about emerging challenges. Standards specific to AI in psychology are still developing, and psychologists have opportunities to shape them.

Consider the environmental and workforce implications of AI adoption. The WHO guidance emphasizes sustainability (WHO, 2021), including environmental impact and workforce disruption. Psychologists making decisions about AI tools should consider these broader implications, though the practical mechanisms for doing so remain underdeveloped.

Engage with the ethics of AI in continuing education. Make AI ethics a routine part of your continuing professional development. The pace of change in this area is rapid, and one-time training is unlikely to be sufficient.

Case Scenario 6: The Practice Group

Dr. B is a senior psychologist in a group practice of eight clinicians. The practice is considering adopting an AI-based platform that would provide symptom tracking between sessions, generate automated session summaries, and offer treatment recommendations to clinicians. The platform's marketing materials are compelling, but Dr. B has concerns about the evidence base and about data practices. The other clinicians in the practice have varying levels of familiarity with AI and varying degrees of enthusiasm for adoption.

Analysis. This scenario provides an opportunity to apply several of the practice recommendations described above. Dr. B might begin by leading a structured evaluation of the platform, including review of published evidence, examination of data practices, and consultation with colleagues who have used it. The group might establish criteria for adoption based on the WHO principles: does the platform have adequate evidence of safety and efficacy for the populations served (well-being and safety); can its role in care be communicated transparently to patients (transparency); does it perform equitably across patient groups (equity); can the practice maintain clinical judgment and accountability in its use (responsibility)? If the platform meets these criteria, adoption may proceed with appropriate monitoring; if not, the group has a principled basis for declining or negotiating changes.

Future Directions and Research Gaps

The Evolving Landscape

The WHO guidance itself acknowledges that "AI for health is a fast-moving, evolving field, and many applications, not yet envisaged, will emerge with ever-greater public and private investment" (WHO, 2021). This observation applies with particular force to mental health care, where new AI applications are emerging rapidly and where the evidence base is often lagging behind adoption. Several areas warrant particular attention going forward.

Empirical Evidence on AI in Mental Health

The empirical literature on AI in mental health care is growing but remains uneven. Well-designed studies of specific tools in specific populations are relatively few, and much of the available evidence comes from developers whose independence is limited. There is a need for more rigorous, independent evaluation of AI tools across the full continuum of mental health care. This includes head-to-head comparisons with traditional interventions, examination of performance across demographic groups, long-term outcome studies, and analysis of unintended consequences.

Psychologists trained in research methods have important contributions to make in this area. The profession's methodological traditions—including attention to validity, cultural sensitivity, and the complexity of human behavior—can strengthen the empirical foundation of AI in mental health.

The Development of Practice Guidelines

Professional practice guidelines for the use of AI in psychology are still nascent. Existing ethical codes provide general principles that apply to AI as they do to other tools, but specific guidance tailored to AI-related challenges is limited. The development of such guidance is an important task for professional associations. It should draw on the WHO framework and other international resources while adapting them to the specific contexts of psychological practice.

Governance and Regulation

The WHO guidance identifies governance as a critical area for further development (WHO, 2021). Existing regulatory frameworks were designed for a pre-AI era and often struggle to address the distinctive features of AI systems, including their adaptivity, opacity, and cross-jurisdictional deployment. The development of appropriate regulatory frameworks is an ongoing challenge that involves multiple actors and multiple levels of government. Psychologists have opportunities to contribute to this development through professional associations, expert testimony, and public engagement.

Training and Workforce Development

The training of psychologists needs to evolve to prepare the next generation for practice in AI-augmented environments. Graduate programs are beginning to include AI-related content, but the depth and consistency of this coverage vary widely. There is a need for curricular development that ensures all psychologists graduate with the critical AI literacy needed for contemporary practice. Continuing education for practicing psychologists is equally important, given the pace of change in this field.

Cross-Disciplinary Collaboration

Effective engagement with AI in mental health requires collaboration across disciplines. Psychologists cannot address these issues alone; they need to work with computer scientists, ethicists, health services researchers, patient advocates, and policymakers. Building the infrastructure for such collaboration—shared conferences, cross-disciplinary journals, joint research programs—is an important task for the field.

Attention to Global Equity

The WHO framing of these issues is inherently global, and the guidance's emphasis on the digital divide (WHO, 2021) is particularly important. Much of the current development of AI in mental health is concentrated in a small number of high-income countries, and there is a risk that tools developed in these contexts will be deployed in low- and middle-income settings without adequate adaptation. Ensuring that AI in mental health serves the world's populations equitably requires deliberate effort, investment, and attention from the psychology profession as a whole.

Unresolved Ethical Tensions

Several ethical tensions raised in the WHO guidance remain unresolved and will require ongoing engagement. These include the tension between innovation and precaution, the balance between individual autonomy and public health goals, the relationship between commercial and public interests in AI development, and the appropriate role of AI in high-stakes clinical decisions. These tensions do not admit of final resolution but will need to be navigated through ongoing dialogue, deliberation, and adaptation.

Conclusion

The integration of artificial intelligence into health care is one of the most significant developments in the contemporary practice of psychology. It brings genuine promise—of expanded access, improved precision, and new insights into human behavior—alongside genuine risks to autonomy, equity, privacy, and the integrity of professional practice. The World Health Organization's 2021 guidance provides a principled framework for navigating this landscape, articulating six ethical principles that together offer a coherent vision of AI that serves human flourishing rather than displacing it.

For psychologists, engaging with AI is not optional. The tools of AI are already present in the health systems in which psychologists practice, in the consumer applications their patients use, and in the broader social environment that shapes psychological experience. The question is not whether psychologists will encounter AI but how they will encounter it. Will they engage as passive consumers of tools developed by others, or as active participants in the ethical shaping of AI in mental health? Will they defer to institutional and commercial pressures, or will they exercise the professional judgment and advocacy that patients need? Will they attend narrowly to the technical features of specific tools, or will they engage with the broader questions of governance, equity, and the meaning of psychological practice in an AI-augmented world?

The WHO guidance offers a starting point, but it is only a starting point. Its principles must be interpreted and applied in specific contexts. Its ethical vision must be operationalized through professional practice, institutional design, and public policy. Its emphasis on human warranty—on the continued centrality of human judgment in health-care decisions—must be sustained against pressures toward automation and abdication.

Psychologists are well positioned to contribute to this work. Their training equips them to think about human experience with the complexity it deserves. Their ethical traditions emphasize respect for persons, attention to power differentials, and commitment to social justice. Their professional identity is grounded in the relationship between clinician and patient, a relationship that AI can support but should not supplant. Bringing these resources to bear on the challenges of AI in health care is one of the important professional tasks of the coming decades. This manual has aimed to support psychologists in taking up that task with the seriousness, sophistication, and ethical grounding that the moment requires.

The stakes are high. The choices being made now—by technology companies, by governments, by health systems, and by individual clinicians—will shape the practice of psychology for generations. If those choices are made without adequate attention to ethics, the profession and the people it serves will bear the consequences. If they are made with the ethical care that the WHO guidance commends, the promise of AI can be realized in ways that genuinely advance human well-being. Psychologists have a distinctive contribution to make to ensuring that the latter path is the one that is taken.

References

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance. Executive summary*. Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.
<https://creativecommons.org/licenses/by-nc-sa/3.0/igo>