Deloitte.

Global governance toolkit for
digital mental health
Building trust in disruptive
technology

April 2021

Contents

Foreword

Our ambition for mental health is a world where every person realizes their desired emotional, psychological, and social potential.

Dear colleagues:

Technology has already transformed our finance, transport, tourism, education, and media industries beyond recognition compared to 20 or even 10 years ago, and it is set to transform the mental health sector as well.[1]

Disruptive technology, such as artificial intelligence and machine learning (AI/ML), digital reality (DR), blockchain, and cloud computing, provide an opportunity to make breakthrough solutions that improve mental health and well-being outcomes on a greater scale than ever before. There are now over 10,000 apps related to mental health in the Apple and Google play stores.[2] Online services such as internet cognitive behavioral therapy (iCBT) courses are becoming more common. Even computer games, wearables, and augmented reality are being showcased for their utility in mental and behavioral health. The overwhelming majority of those 10,000 apps, it should be noted, are not currently evidence based.[3]

Using disruptive technologies in digital mental health solutions raises new ethical questions about safety, efficacy, equity, and sustainability—many of which are not answered by current regulatory structures and leave stakeholders open to possible privacy, quality, and safety risks concerns. With that in mind, the World Economic Forum and Deloitte have developed a toolkit that aims to provide governments, regulators and independent assurance bodies with the tools to protect personal data, ensure high quality of service and address safety concerns. This toolkit is a practical next step to help government, regulators, health care and insurance organizations, and service vendors:

- Understand the potential for digital mental health services in improving the mental and behavioral health of all people;

- Develop principles and standards for the safe, ethical, and strategic implementation of digital mental health services;

- Adapt, pilot, and adopt these standards and principles in countries, jurisdictions, health systems, and across the globe;

- Improve access, efficacy, quality, and safety of digital mental health solutions by adopting better practices and standards;

- Make strategic investment and incentivization decisions in the global digital mental health ecosystem to encourage its growth; and

- Make informed decisions to seamlessly incorporate digital mental health tools into a health system, workplace, community, product, or service.

If you have questions or want to have a deeper discussion, please reach out to me or the individuals named at the conclusion of this report.

Sincerely,
Dr. Stephanie Allen
Deloitte Global Health Care Leader
steallen@deloitte.com.au

Introduction

Between a quarter and half of the global population is affected by a mental disorder at some point in their lives.[4]

Mental health and behavioral health disorders place a significant and under-recognized burden on our global health care system, on people's social needs, on basic human rights, and on the economy. The human cost is immense: 800,000 people commit suicide every year, and suicide is the leading cause of death among young people. Individuals suffering from untreated mental health disorders are unable to realize their desired potential and all too often are exposed to a wide range of human rights violations.[5] Between 2011 and 2030, the cumulative economic output loss associated with mental disorders is projected to be US$ 16.3 trillion worldwide.[6] Secondary consequences of mental ill-health are estimated to cost employers US$ 2,000 per employee per year from presenteeism, absenteeism, and unnecessary turnover.[7] The direct and indirect costs can amount to five percent of a country's gross domestic product (GDP).

Recent health, climate, and political crises, such as the COVID-19 pandemic, have highlighted and exacerbated the world's mental health challenges. Public health agencies have warned that a wave of depression, suicide, and other mental ill-health issues are on the horizon due to multiple crises in 2020.[8,9] COVID-19 has also accelerated digital change at an unprecedented rate, transforming services and offering a huge opportunity to move forward with changes in our mental health systems.

Disruptive technologies have great potential to transform global mental health and behavioral health systems to be more fit-for-purpose, more affordable, and more easily scaled, particularly for countries and people without adequate access to mental health services. However, using disruptive technologies in digital mental health solutions raises new ethical questions about safety, efficacy, equity, and sustainability. Many of these questions are not answered by current regulatory structures and leave us open to possible privacy, quality, and safety risk concerns.

This toolkit provides government, regulators, and independent assurance bodies the tools to develop, adopt, and engage standards and policies that address major ethical concerns, thereby protecting consumers, enabling them to more easily assess quality mental health care and make informed choices about their own mental health, and encouraging the strategic growth of safe, ethical, and effective digital mental health services. The toolkit also provides government, businesses, and service vendors a guide by which they may operate ethically and safely, for the betterment of consumers and the ecosystem in which they intend to operate. It does this by exploring the current and future opportunities and ethical issues presented by disruptive technologies in mental health, and offering a framework of governance principles, standards, and processes.

Who will find this toolkit useful and why?

Government and regulators

The primary goal of the toolkit is to influence government health departments and health regulators to pilot and adopt principles and standards that encourage the safe, ethical, and strategic implementation of digital mental health services in their respective jurisdictions. The potential impact is the adoption of services that offer low-, middle-, and high-income populations access to scalable, effective, affordable, and needs-based services. The toolkit also provides foundational insights to aid strategic decision-making vital to understanding the levers that are available to incentivize growth, make more strategic commissioning decisions, and augment and integrate with existing health care models for a more holistic, seamless, mental and behavioral health system.

Health care and insurance organizations

This toolkit provides health care and insurance organizations rubrics to understand digital mental health services and the core components to be considered in order to assess, build, and integrate high-quality digital mental health services. The principles are fundamental to making effective commissioning and investment decisions, creating effective combined models of care, and encouraging safe innovation in approaches to consumers' mental and behavioral health.
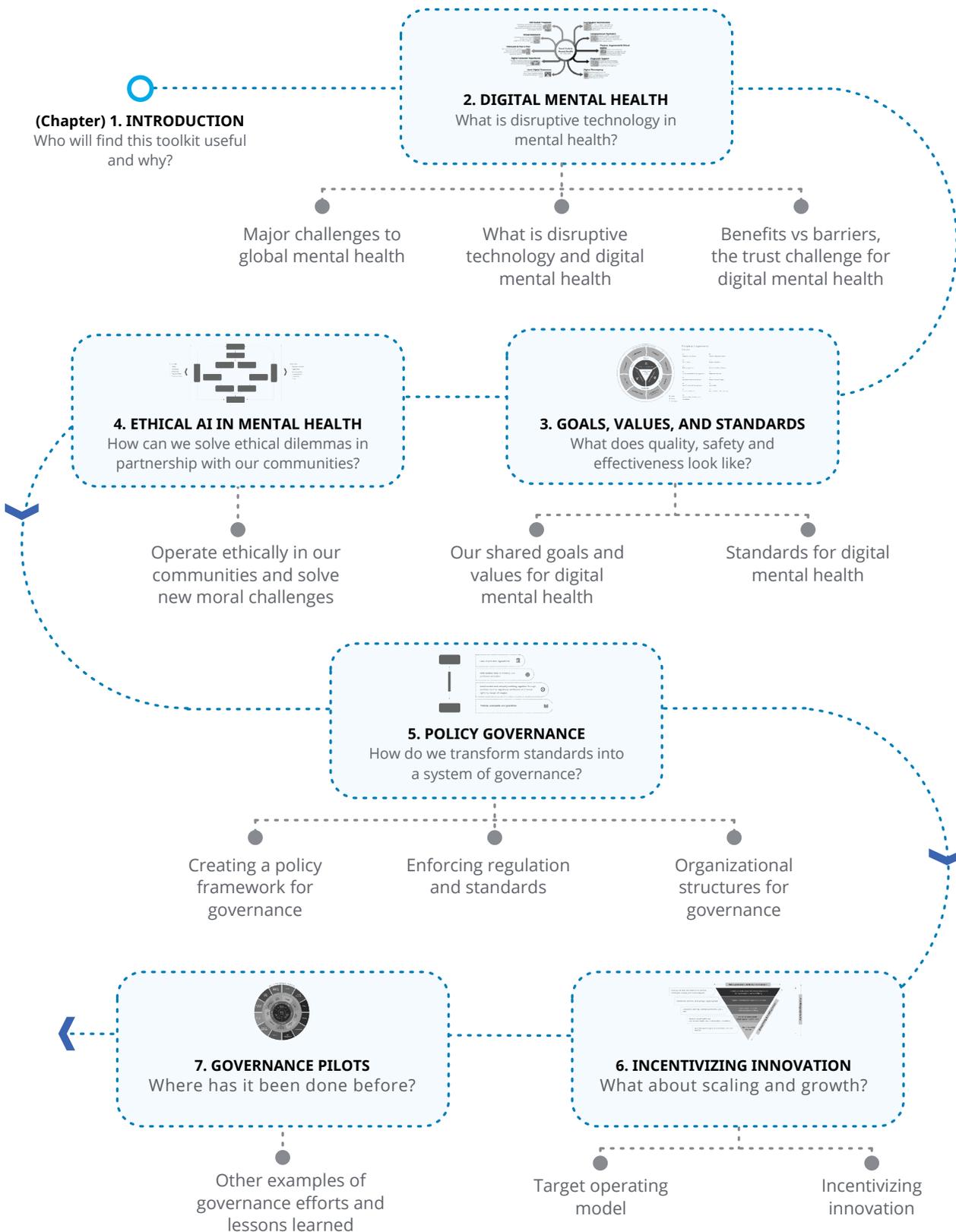
Digital mental health innovators and vendors

By bringing together the views of consumers, clinicians, innovators, government, and health care providers, this toolkit offers a roadmap for digital mental health innovators and vendors to create trusted services. The standards detail the potential critical interests of those stakeholders. When these standards are applied to their services, they are the means to exceed consumers' expectations and accelerate opportunities for adoption, investment, and integration.

This toolkit enables health care ecosystem stakeholders to embrace validated digital mental health services safely, strategically, and ethically by helping to foster consumer trust in digital mental health services through assurance and transparency. Doing so should support greater consumer engagement and better outcomes. Also, if clinicians trust these services, they are more likely to integrate use of these services into their practice, potentially increasing their reach, affordability, and effectiveness.

Navigating the toolkit

(Chapter) 1. INTRODUCTION
Who will find this toolkit useful
and why?

2. DIGITAL MENTAL HEALTH
What is disruptive technology in
mental health?

Major challenges to
global mental health

What is disruptive
technology and digital
mental health

Benefits vs barriers,
the trust challenge for
digital mental health

4. ETHICAL AI IN MENTAL HEALTH
How can we solve ethical dilemmas in
partnership with our communities?

3. GOALS, VALUES, AND STANDARDS
What does quality, safety and
effectiveness look like?

Operate ethically in our
communities and solve
new moral challenges

Our shared goals and
values for digital
mental health

Standards for digital
mental health

5. POLICY GOVERNANCE
How do we transform standards into
a system of governance?

Creating a policy
framework for
governance

Enforcing regulation
and standards

Organizational
structures for
governance

7. GOVERNANCE PILOTS
Where has it been done before?

6. INCENTIVIZING INNOVATION
What about scaling and growth?

Other examples of
governance efforts and
lessons learned

Target operating
model

Incentivizing
innovation

Digital mental health

Disruptive technologies provide an opportunity to overcome challenges and create breakthrough solutions that improve mental health and well-being outcomes on a greater scale than ever before. Still, ethical concerns are being raised about these new tools and services.

Although some governments, care providers, insurers, and employers are starting to invest in addressing the mental health crisis, it is generally recognized by The Lancet, World Health Organization (WHO) and many others that there are significant global and local barriers to positive change in mental health (figure 1).
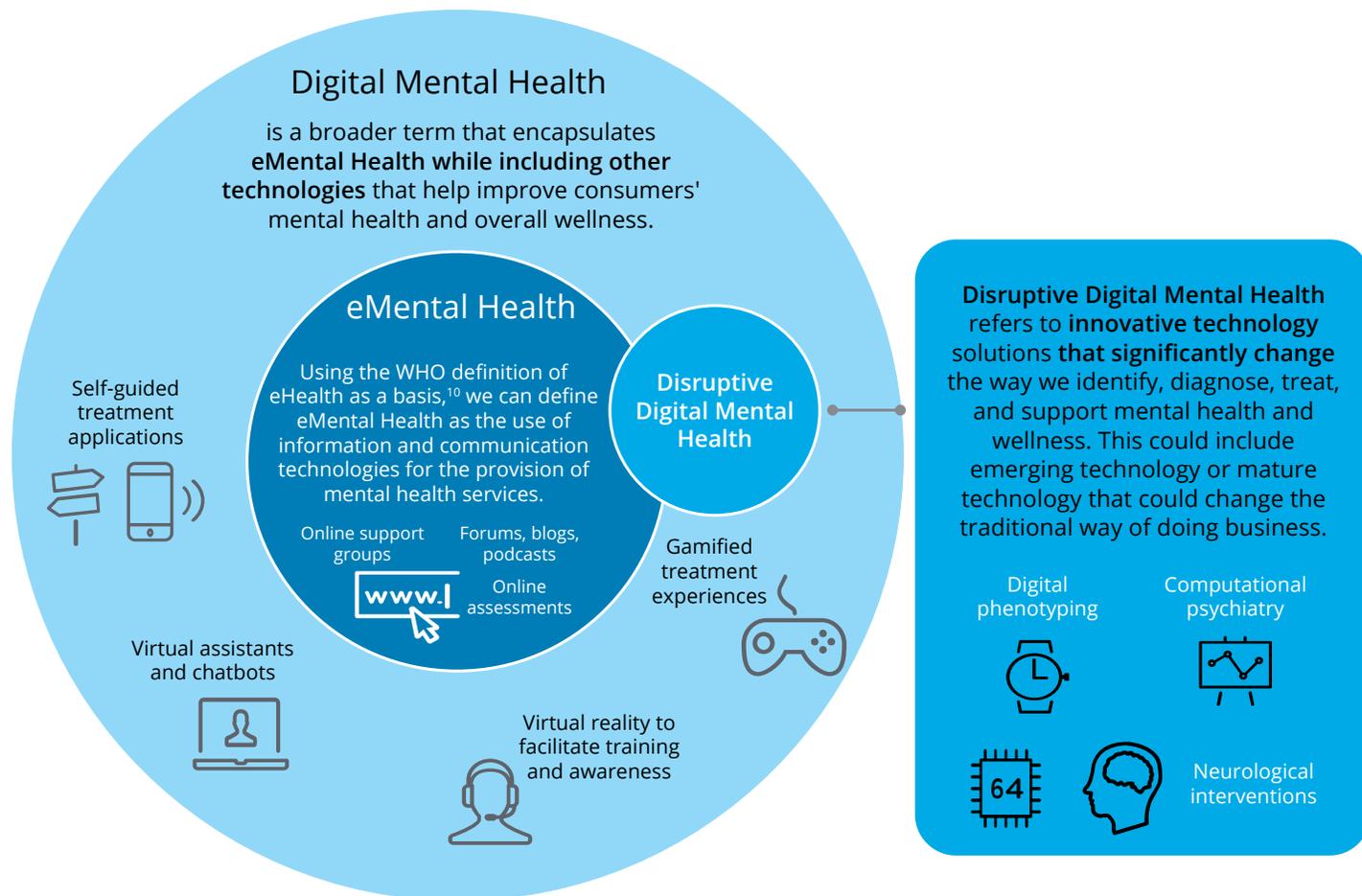
Figure 1. Major challenges faced in achieving better global mental health outcomes

Stigma, discrimination, and absence of legal and medical protections	Chronic underspend, inaction, and lack of access	Sub-scale and siloed health care systems	Gaps in clinical and scientific knowledge	Changing determinants of mental health

The role of disruptive technology in mental health

Disruptive technology in mental health refers to innovative technology-based solutions that significantly change the way we prevent, identify, diagnose, treat, and support mental health and wellness (figure 2). Offerings are broad and include both emerging technologies and mature Digital Mental Health and eMental Health solutions.
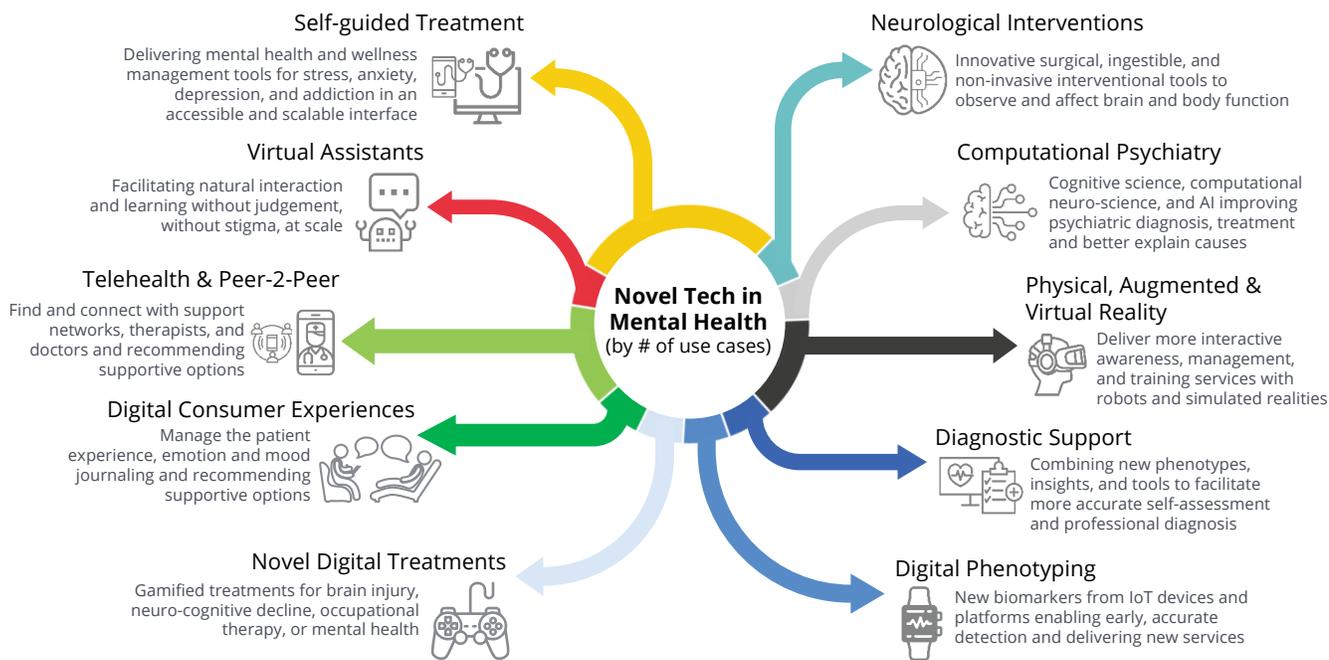
Figure 2. The role of disruptive technologies in mental and digital mental health

Digital Mental Health

is a broader term that encapsulates **eMental Health while including other technologies** that help improve consumers' mental health and overall wellness.

Self-guided treatment applications

eMental Health

Using the WHO definition of eHealth as a basis,[10] we can define eMental Health as the use of information and communication technologies for the provision of mental health services.

Online support groups

Forums, blogs, podcasts

www.I

Online assessments

Disruptive Digital Mental Health

Gamified treatment experiences

Virtual assistants and chatbots

Virtual reality to facilitate training and awareness

Disruptive Digital Mental Health refers to **innovative technology** solutions **that significantly change** the way we identify, diagnose, treat, and support mental health and wellness. This could include emerging technology or mature technology that could change the traditional way of doing business.

Digital phenotyping

Computational psychiatry

Neurological interventions

For our research, we gathered and analyzed 190 use cases of information and computer technology being applied to mental health and well-being services, tools, and solutions where there was evidence of more than 10,000 users. Results show the incredible breadth and diversity in digital tools currently available, ranging from resilience and prevention services to targeted diagnosis and treatment tools; from basic educational websites to virtual reality avatars delivering cognitive behavioral therapy; from Excel spreadsheets guiding consumers in a local community to effective services to globe-spanning electronic mental health records that support integration of services and computational psychiatric research efforts; and from simple phone applications developed by small agile teams to implantable neurological nano devices that affect addiction and brainwave measuring headsets—delivered at consumer-appropriate prices—to optimize personal stress and relaxation for enhanced well-being and productivity.
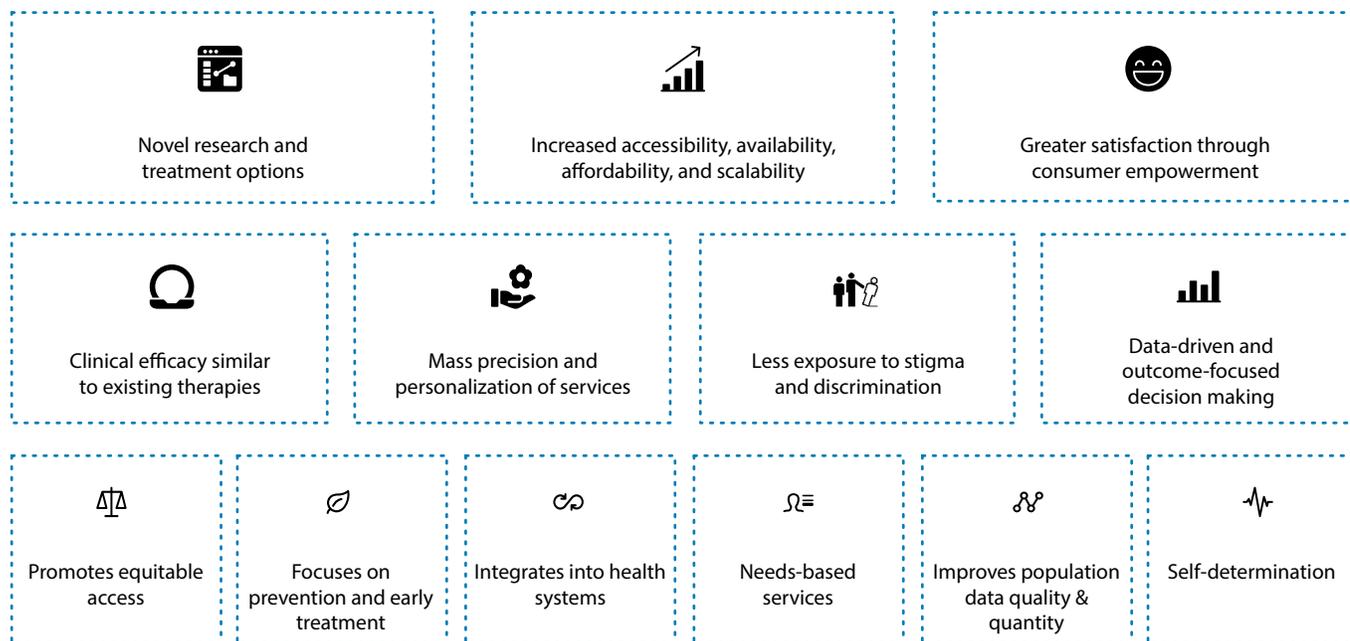
Figure 3 highlights some of the categories of innovation identified in our 190 use cases, with the proportion in each category indicated around the circle.

Figure 3. Novel technologies in mental health

Self-guided Treatment
Delivering mental health and wellness management tools for stress, anxiety, depression, and addiction in an accessible and scalable interface

Virtual Assistants
Facilitating natural interaction and learning without judgement, without stigma, at scale

Telehealth & Peer-2-Peer
Find and connect with support networks, therapists, and doctors and recommending supportive options

Digital Consumer Experiences
Manage the patient experience, emotion and mood journaling and recommending supportive options

Novel Digital Treatments
Gamified treatments for brain injury, neuro-cognitive decline, occupational therapy, or mental health

Novel Tech in Mental Health
(by # of use cases)

Neurological Interventions
Innovative surgical, ingestible, and non-invasive interventional tools to observe and affect brain and body function

Computational Psychiatry
Cognitive science, computational neuro-science, and AI improving psychiatric diagnosis, treatment and better explain causes

Physical, Augmented & Virtual Reality
Deliver more interactive awareness, management, and training services with robots and simulated realities

Diagnostic Support
Combining new phenotypes, insights, and tools to facilitate more accurate self-assessment and professional diagnosis

Digital Phenotyping
New biomarkers from IoT devices and platforms enabling early, accurate detection and delivering new services

Digital mental health services offer a range of potential direct and indirect benefits that promise to augment, supplement, and scale existing mental health models of care (figure 4). These tools are most effective when they are either used in combination with current care models or to access populations that can't or don't currently access traditional medical services.

Figure 4. Summary of the potential benefits of digital mental health

Novel research and treatment options	Increased accessibility, availability, affordability, and scalability	Greater satisfaction through consumer empowerment

Clinical efficacy similar to existing therapies	Mass precision and personalization of services	Less exposure to stigma and discrimination	Data-driven and outcome-focused decision making

Promotes equitable access	Focuses on prevention and early treatment	Integrates into health systems	Needs-based services	Improves population data quality & quantity	Self-determination

Trust, evidence, and data challenges for digital mental health

Despite digital mental health's potential, our interviews with consumers, clinicians, and health care system owners often reveal safety and efficacy concerns, as well as ethical questions related to use of data and AI/ML in mental health services. Trust in a mental health service is fundamental to its uptake, therapeutic outcomes, and sustainability.

There are two main factors for consumers, clinicians, and health system owners to consider when evaluating digital mental health services/tools. First, the technology must be safe for users and effective in leading clinical improvement. Second, the technology must be trustworthy: Consumers trust that the tool is using proven, effective means to achieve its advertised outcomes, that their information is protected and not used in a harmful manner, and that the tool has been designed to uphold their best interests.

Several ethical risks with potential for harm may act as barriers to the adoption of disruptive technologies in mental health care and to high-quality consumer outcomes. An in-depth understanding of these risks is important to developing technology that can be used and trusted. Through interviews with leading innovators and health care professionals and current literature reviews, we have categorized these risks into six core areas:

1. Untested efficacy or misleading claims of mental health benefits

Making unsubstantiated or misleading claims about the health benefits of mental health services can cause unintended harm to the consumer (case study 1). For example, a digital product may use aspects of clinical treatment that have not been trialed or validated in that product's delivery mode. While it is not necessarily a vendor's intention to provide unsubstantiated or misleading claims about a product's health benefits, misalignment between advertised claims and services or failure to mention associated risks or conditions that should be precluded may place consumers at risk of unknowingly engaging with treatment that unintentionally harms them or worsens their mental condition.

Case study 1

Proactive mental health intervention by social media company[11]

A social media start-up offers users a popular feature to live stream moments in their life. Some users elect to use the live stream feature to share their self-harm experiences, food restriction strategies for people with anorexic behaviors, and even some suicide events. After several such cases received negative publicity, the company launched an internal program to screen for self-harm behaviors. Anything flagged by the AI is reviewed by a trained staff member. The program has not been reviewed by a health regulatory agency but, according to the start-up's internal research team, may have saved 450 lives in the first two months alone.

1. What is the responsibility of the social media company to address these cases? Is the company liable for harm to the user if they don't report the prediction to police or emergency help?

2. What kind of consent is needed? Should users be allowed to opt-out or should users only opt-in to this function? What level of privacy can be promised to users, given the possibility that the algorithm may alert friends, family, and emergency authorities?

3. If the company has an internal AI program that predicts self-harm, what is the optimal accuracy for these predictions? What are best practices for training staff members to evaluate the AI prediction? What are the harms of an incorrect prediction?

4. What are best practices for programming, such as AI that would address biases that could exacerbate existing disparities?

2. Health data privacy breaches

The collection, storage, and use of individuals' mental health data poses several ethical risks. Video and audio data pose a particularly high risk for exposing sensitive and personal information due to the underdeveloped technical and legal privacy protections that exist to protect this type of data. Consumers are particularly concerned that their private information will be used by insurance companies to deny or affect their coverage, for punitive reasons by government agencies, or sold to third parties and used for non-therapeutic purposes. Currently, frameworks to protect such data are not widely used and should be further developed to ensure optimal efficacy.[12]

3. **Harm to patient through malfunction, incorrect advice, or misuse**

A digital mental health service that malfunctions, is misused, or misinforms consumers can negatively impact their experience and health. Most pertinently, bots, such as those being trialed as chatbots in mental health apps, may malfunction in unpredictable ways, which could harm a consumer. Without proper oversight, a digital service may provide information and advice that has not been validated either by clinical evidence or experienced professionals (case study 1). Services also may provide advice that is not matched to a consumer's specific social, cultural, or relationship situation, resulting in poor or negative outcomes. Finally, without oversight, a model that is designed to use psychologically influential techniques to modify a consumer's behavior and thought processes may be used to affect political, abusive, criminal, and other non-beneficial outcomes.

4. **Lack of accountability or incentives for prioritizing consumers' best interest**

Technology regulation in mental health care is currently limited: Product and services vendors are not held accountable to an appointed or competent regulatory body or to an overarching set of standards, placing consumers' well-being at risk (case study 2). This issue is amplified by the cross-border nature of digital services. Companies may train their ML models to achieve outcomes that are not in the best interest of their consumers' demographic, such as increasing the product/service's addictiveness and profitability, rather than aim to improve the therapeutic outcome.

Case study 2

The pandemic stress app with 5-star reviews[13]

An app claims to improve mood and reduce stress during COVID isolation and, secondarily, boost the immune system. The app store profile has 45,000 5-star and 4-star reviews. The press has celebrated the app's developer as a brilliant engineering student with no medical background who is on a mission to increase access to mental health for all. There are no publications from any randomized trials of the app, but the app's profile cites 25 publications from other trials (some peer-reviewed, others not) linking stress and depression to increased risk for viral infections. Recent peer review studies have also demonstrated that some placebo versions of popular mindfulness apps (like symptom trackers) resulted in gains similar with the app; a digital placebo effect. This placebo effect further complicates the evidence needed to regulate the efficacy of any (medical) apps.

1. What are the risks and harms of an app that offers digital placebo relief versus that of an app considered by scientists/regulators to be medically efficacious? Is the digital placebo effect more helpful than having no access to mental health providers?

2. How can mental health providers, researchers, and advocates help the public distinguish between effective and ineffective apps or high product performance versus low performance?

3. Some app developers have stated that all users should own their own data. However, the data often becomes more valuable than the app itself. How can independent, small app developers be incentivized to help protect user data?

5. Inaccessible or inequitable design that widens the digital divide

Technology can increase the global inequality and inaccessibility of mental health care in two ways. First, access to a digital device which can provide mental health care is not equitable. For example, 184 million fewer women own mobile phones compared with men.[14] Additionally, there is a significant technology literacy gap in both consumers and clinicians worldwide, which limits consumers' ability to access digital mental health care and diminishes clinicians' capability to deliver it.[15] Such discrepancies have the potential to widen the gap for mental health care access, particularly for minorities and those of low socioeconomic status.

Second, AI-enabled mental health devices may develop bias as a result of the way they are trained and tested, which may harm consumers in unintended ways.[16] One form of algorithmic bias occurs when models are trained on a dataset that does not reflect the demographic of people for whom it will be used. Additionally, unsupervised models may develop bias by associating certain features, events, or words with certain demographics or socioeconomic groups as it learns on a dataset, especially if the dataset is seeded with biased labeling.[17] As such, integrating AI into mental health care poses the risk of harming or detracting from the care of disadvantaged groups of people.

6. Long-term social and behavioral consequences

The implementation of technology in mental health care poses the risk of impacting consumers' long-term behavioral and social capabilities. Worsening mental health symptoms have been associated with prolonged screen time, particularly in children and adolescents, which may highlight an unintended risk of integrating technology into mental health care.[18] Additionally, recent publications have posited that technology developed for "self-care" or self-treatment of mental health lacks the therapeutic relationship that a consumer would have with a health care professional, and may exacerbate the behavioral traits of a person experiencing issues with mental health, including social withdrawal and isolation.[19,20] Research also has identified that disclosing personal health information online, where it may be accessed by relations, employers or others, may have unintended social and behavioral consequences.[21]

Additional ethical domains for consideration

Current ethical guardrails for individuals and organizations rendering scientific, educational, or professional services in psychology, psychiatry, and mental health include, among others, the Code of Ethics governed and administered by the American Psychological Association; standards of conduct, performance, and ethics administered by the Health and Care Professions Council (HCPC) and Software as a Medical Device (SaMD); and decisions by the US Food & Drug Administration (FDA). These guardrails are not necessarily appropriate for tackling the unique ethical challenges of data and AI in psychology and mental health, as they were designed when these risks did not exist. Further, these guardrails do not always cover the range of services that are blurring the lines between consumer products, self-improvement products, and mental health treatments. Considering the primary risks and central principles that are recognized as important for digital mental health services, as discussed in the next chapter, these guardrails could potentially be strengthened by considering four ethical domains:

- **Health data protection:** Taking responsibility for proper stewardship of people's personal, health, and mental health information;

- **AI/ML in health care:** The ethical application of AI/ML in medical, psychiatric, and psychological fields;

- **Human-centered design:** Providing engaging experiences that are fit for purpose and usable, promoting wider access, adoption, and follow-through; and

- **Responsible business:** Operating ethically, sustainably, and responsibly for the best interests of consumers and society.

Goals, values, and standards

Digital mental health standards should be developed in line with the cultural, medical, and social principles and values of the consumers and clinicians who are using those services and implemented to continuously and flexibly maintain high levels of safety and ethics on behalf of those stakeholders.
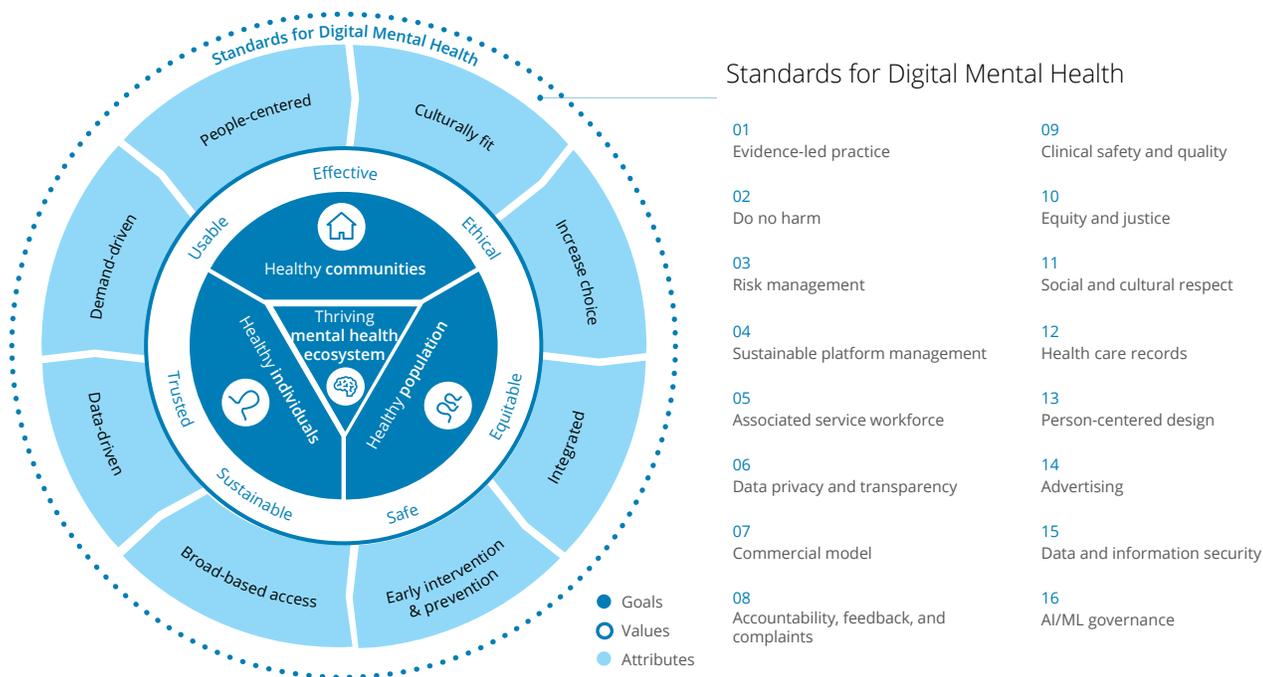
When our goal is to ensure that digital mental health solutions are safe, strategic, and trusted for everyone involved, there are many ways to approach creating and enforcing governance systems creation. This chapter focuses on developing two core methodologies: a standards-based approach to regulation that can be applied widely to ensure a service meets the ethical expectations of its stakeholders, and an ongoing governance model that a business or digital mental health service vendor could employ to resolve new ethical dilemmas that arise from innovation and integrate them within their targeted community.

Standards/policy creation may spring from a variety of sources, among them:

• Identifying the ethical way an organization should behave and establishing the rights and values that it should uphold;

• Identifying specific risks involved for an organization's outcomes and establishing required mitigations; and

• Identifying systems that should be in place for an organization to ensure safe and quality outcomes.

Our methodology for developing robust standards for digital mental health (figure 5) is purpose-driven, outcome-orientated, and flexible. It starts with identifying the framework's goals and developing the shared values and attributes which define the purpose and principles for digital mental health. These are based on: 1) consultations with consumers, innovators, and health systems owners around the world; 2) a research and literature review; and 3) our pilot project developing the Digital Mental Health and Addiction Services Evaluation Framework with the New Zealand Ministry of Health.

Figure 5. Goals, values, and standards for Digital Mental Health

Standards for Digital Mental Health

01 Evidence-led practice	**09** Clinical safety and quality
02 Do no harm	**10** Equity and justice
03 Risk management	**11** Social and cultural respect
04 Sustainable platform management	**12** Health care records
05 Associated service workforce	**13** Person-centered design
06 Data privacy and transparency	**14** Advertising
07 Commercial model	**15** Data and information security
08 Accountability, feedback, and complaints	**16** AI/ML governance

The resulting standards to successfully audit and govern a digital mental health service and provide assurance of its safety, quality, and efficacy are outcome-focused, allowing for greater flexibility in implementation options and future-proofing requirements. The intended outcome of auditing a service against these requirements is a categorized grade where the service is scored according to its ability to meet or exceed various requirements in those categories. Note that flexibility in applying and validating digital standards for mental health is important, as applications vary widely—from simple informational services to software that supports secondary psychiatric diagnosis. In addition, there may be jurisdictionally different views on social, cultural, and medical practices to consider, as well as existing legislation for health devices, health software, data privacy, health information privacy, and responsible business practices, equality, and justice.

The following list summarizes each sub-group within the standards presented in the toolkit.

1. Lead practice with evidence: Utilize clinically validated therapeutic models of care; offer content and therapeutic approaches based on science; support ethical research and publication

2. Do no harm: Promote ethical decision making; incorporate processes to do no harm

3. Establish a risk management culture: Govern responsibly with processes for risk management; govern and manage risks proactively; create a risk management culture

4. Ensure clinical safety and quality: Monitor clinical safety and quality; operate with a defined model of care; manage the clinical safety and quality performance; manage adverse events around clinical safety and quality; continuously improve clinical safety and quality; prevent harm and suicide

5. Commit to a robust service workforce: Operate with an appropriately qualified service workforce; ensure clarity in roles and responsibilities; ensure the service workforce is skilled; manage service workforce performance; ensure cultural and safety training for the service workforce

6. Promote data privacy and transparency: Use informed consent for data privacy; have system controls for data privacy; allow data management by the consumer; gain consent to capture data; gain informed consent for data usage

7. Maintain data and information security: Have secure data and information through systems for data encryption; have secure data and information through systems for data security; have secure data and information through systems for consumer authentication

8. Ensure health care records interoperability: Make responsible use of health care records; provide interoperable health care records; ensure accuracy of health care records; enable bi-directional management of care; manage referrals responsibly

9. Orient around person-centered design: Promote accessibility and reach; ensure high usability; promote positive user experience; tailor communications and services to diverse groups; co-design service and governance with service consumers; empower service consumers to make decisions relevant to their care; use a person-centered model of care

10. Promote social and cultural respect: Ensure services are free from discrimination; show social and cultural respect

11. Commit to equity and justice: Promote equitable health outcomes; prevent racism and marginalization; prevent coercion or exploitation in services; govern responsibly on ethical principles or values; govern responsibly on equity, ethics, and diversity

12. Advertise responsibly and accurately: Make responsible therapeutic claims; responsibly advertise by avoiding false and deceptive statements; ensure responsible and ethical in-product advertising

13. Enact a responsible business model: Operate a financially viable business model; put in place a responsible business model; be transparent about costs to service consumers

14. Operate with accountability and accept feedback and complaints: Demonstrate accountability for the tool; uphold the rights of service consumers

15. Provide a sustainable platform and continuity of care: Keep systems and services up to date; ensure platform stability and availability

16. Govern AI/ML responsibly: Ensure quality systems and good AI/ML practices; conduct pre-release review and modification governance; monitor and publish real-world performance

Ethical AI in mental health

Our approach to ethical and moral dilemmas creates a social license to operate with our communities and acts as a bridge to ethical innovation in mental health.

A framework for the ethical operation of disruptive technology and AI in mental health should enable transparency, fairness, social capital, and a way forward for service vendors in an open universe of ethical situations (figure 6).

Figure 6. Framework for the ethical operation of AI in mental health
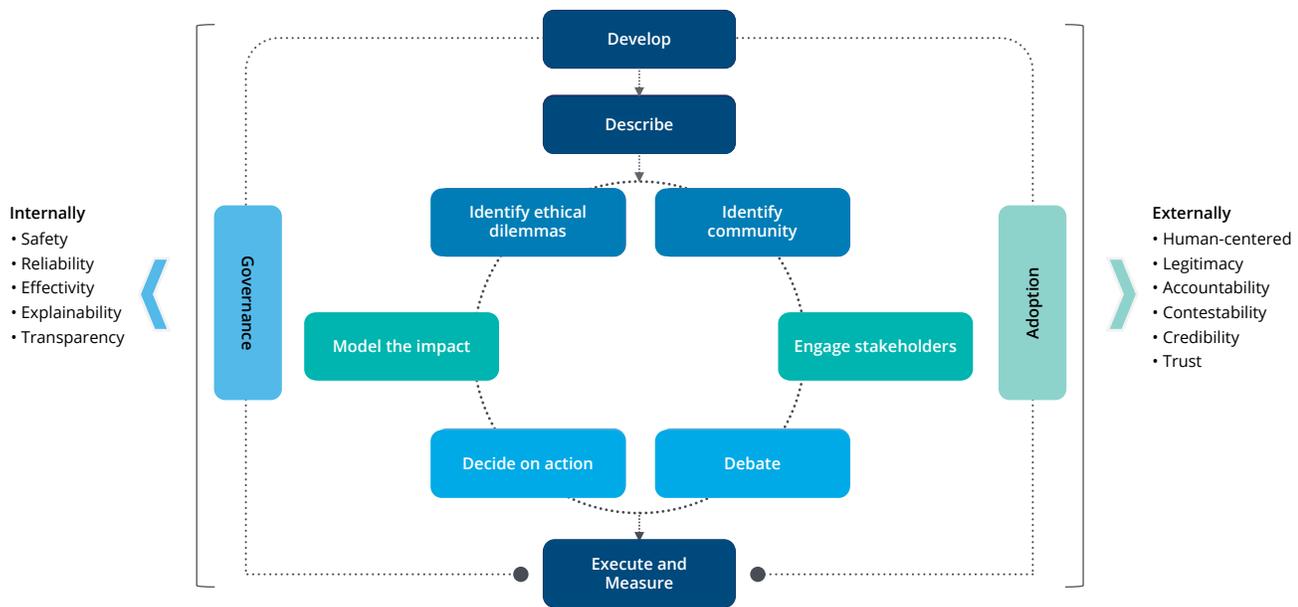
Develop and accurately describe the offering's functional intent and identify targeted stakeholder communities. An accurate description of the tool, service, or decision AI's functional intent provides a basis for identifying risks and ethical dilemmas, and a starting point to engage the community in solving these dilemmas. The description may include targeted consumers and stakeholders, intended state of consumers' health situation or conditions, significance of the service to managing those health conditions, information and data, and decisions and actions.

Identify the value conflicts that may generate ethical dilemmas emanating from effects within the community. An ethical dilemma may come from uncontrolled bias within the dataset, inappropriate choice of algorithm or outcome measures, or the feedback loop that is introduced into an ecosystem with a new service or AI. Two questions that may be asked when assessing a dilemma are: what are the objectives or rights that we are balancing in this situation, and what is the scale of worst-case to best-case scenario in this situation?

Model the clinical impact. Measuring the impact to the consumer and the environment of introducing a system is fundamental to assessing whether the intervention is appropriate compared to the possible damage. Measuring the impact also brings clarity and transparency to the scale of the intervention and effect. Clinical outcome measures may be based on symptomatology (e.g., CORE-OM, OQ-45), social disability (e.g., HONOS), needs (CAN), service usage or, increasingly, recovery outcomes (CHIME). Measuring must also be considered for non-clinical, but still important, factors such as environmental, financial, and social impacts.

Engage stakeholders. To gain a social or moral license to operate, an organization must engage with the affected community to identify and explore the impossible, acceptable, and unacceptable solutions for that community.

Debate and decide on action. A chosen course of action in an ethical situation may not be the consensus view of the community, yet it is informed by what the community may view as acceptable and unacceptable choices. An action may have intended and unintended outcomes, and as such should be continuously monitored.

Execute, measure, and iterate. Modeling the impact of actions taken must continue after implementation to confirm assumptions, review consequences, and unearth unintended outcomes. Doing so also provides transparency and builds trust with the community.

Ensure continuous evaluation. The US FDA's Proposed Regulatory Framework for Modifications to Artificial intelligence/Machine Learning-Based Software as a Medical Device[22] provides a roadmap for continual AI/ML governance:

- An organizational culture of quality and excellence with quality systems and good machine-learning practices;

- Pre-market review and specification of safety, ethics, and effectiveness;

- Algorithm change protocols to control the risks as expected performance or input modifications are introduced; and

- Real-world performance monitoring and publishing of results.

Policy governance

Functional, robust, and flexible policy governance will incentivize the digital mental health ecosystem to be safe, strategic, and ethical, for the good of consumers and service providers.

Several key elements comprise an effective digital mental health evaluation framework (figure 7).

Figure 7. Digital mental health evaluation framework system

1. Framework vision and purpose

2. Framework stakeholders' values and objectives

3. Framework services scope

4. Service description

5. Service classification and risk assessment

6. Service assessment criteria

- Evidence-led practice
- Do no harm
- Risk management
- Clinical safety and quality
- Associated service workforce
- Sustainable platform management
- Date privacy and transparency
- Data and information security
- Healthcare records interoperability
- Person-centered design
- Social and cultural respect
- Accessibility, equity and justice
- Advertising
- Governance and commercial model
- Accountability, feedback, and complaints

7. Service scoring and publishing

Governance

1. **Framework vision and purpose**. Clearly articulating the assessment framework's vision and purpose is a way to drive overarching goals and share them with health care system stakeholders. In this instance, the assessment framework should encourage products that are safe, trusted, effective, equitable, ethical, and sustainable. Also, vision and purpose should be refocused by the regulatory body in their regulatory, clinical, and cultural environment.

2. **Framework stakeholders' values and objectives.** An assessment framework's vision and purpose influences, and is influenced by, its stakeholders' values, principles, and objectives. A human-centered approach should be used to identify these stakeholders and what they want to achieve by implementing the digital mental health regulation. Ideally, the framework would optimize consumers' well-being and outcomes and support development of novel technology in mental health care.

3. **Framework services scope.** In this element, the regulator clarifies the applications and services which lie within the regulation's jurisdiction. Included and excluded services should be determined by the therapeutic mental health and well-being outcomes the services intend to achieve and the use of information and communication technology to support that therapeutic outcome.

4. **Service description.** A vendor's accurate, honest, and easily understood description of the digital mental health service is essential to aid in classifying the service for assessment and provide consumers and clinicians the information they need to make an informed choice for its prescription and use. An accurate description will contain provider information, product description, clinical evidence, data usage and privacy, security, and a business model.

5. **Service classification and risk assessment.** The level of regulations put in place via the assessment framework should reflect the level of risk that the product or service poses to the user. As such, a risk assessment is important to determine the severity of the regulations that the regulator must implement. Classification across therapeutic claims, functional scope, data usage, and workforce should be used to define mandatory, non-mandatory, and supplementary assessment questions.

6. **Service assessment criteria.** The assessment criteria's structure is crucial to provide the vendor with a clear set of requirements to achieve. The criteria should be predicated on the ethical principles that underlie the regulations, thus communicating the importance of compliance. Additionally, the criteria should provide pass and fail examples of how to comply with a requirement in a particular jurisdiction.

7. **Service scoring and publishing.** Linking each assessment criteria to one of the key values of safe, trusted, ethical, effective, equitable, and sustainable performance will produce a scoring rubric which indicates levels of adherence to each of these outcomes and can be used by consumers to quickly assess and compare services.

Regulatory enforcement of digital mental health services can vary along a low- to high-touch continuum. Guidance may range from voluntary standards, kitemarks, and independent body regulations/inclusion in a pre-approved library to more strict pre-market authorization, acts, and laws.

Incentivizing innovation

The aim of a digital mental health governance framework is not to needlessly restrict innovation, but to enable and incentivize the ecosystem to innovate, scale, and thrive with shared goals in mind.

By focusing governance adherence incentives, efforts, and finances on mental health and well-being, there is a huge opportunity to improve both health and economic outcomes. The following is a list of potential incentive mechanisms.

Lever	Description
Financial incentives	Access to funding is still the number-one barrier listed by innovators and providers in the digital mental health space, especially for operating in lower- and middle-income countries. Options include: • Venture capital funding • Grants, foundations, and social bonds • Integration with public health care pricing mechanisms • Integration with public and private insurance pricing mechanisms
Operational education and support	Service innovators and providers can be supported through workshops and education programs on how to improve quality and safety, build capacity for sustainability, and targeting their services based on common issues or strategic goals. Sample programs and relevant topics include: • Health approval and commercialization • Human-centered and co-design methods for mental health • Data, privacy, and security for digital mental health • Ethics of AI in health care and mental health • Cultural and social adaption for vulnerable communities
Research, data, and testing	Service vendors require access to consumers and high-quality data to perform efficacy and effectiveness studies, to advance research, and to understand how to customize their service to meet a population's evolving needs. This may be done through: • Consumer research groups • High-quality, anonymized electronic mental health care records • Anonymized population and demographic data • University and academic research groups, platforms, and networks

Lever	Description
Access to and integration with the market	Access to consumers and the market is often a bi-directional relationship and many mental health consumers will present in existing health systems. As such, integration is vital and may be achieved through: • Existing health and mental health systems as a new option for consumers • Recommendations to clinicians in a pre-approved library, clinical education material, or clinician governance groups
Communities of expertise	Communities of expertise and knowledge storehouses are vital to building on the work of others. These may be initiated or developed through: • Universities • Research hospitals • Symposiums and conferences • Cooperative research centers • Collaborative research bodies
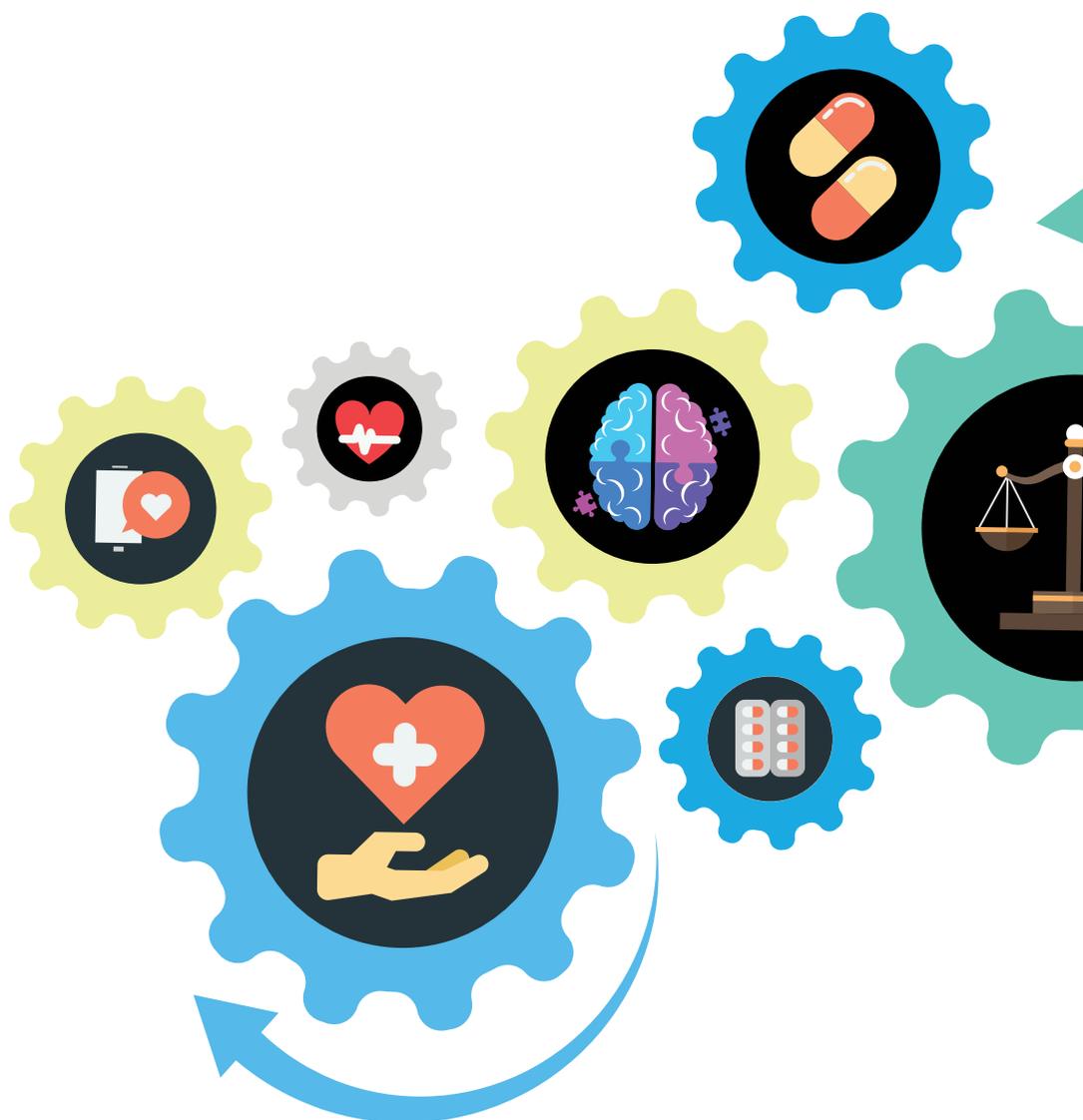
Governance pilots

Pilot lessons from New Zealand

**Digital Mental Health and Addiction Service Evaluation Framework
(DMHAS Evaluation Framework)**
New Zealand Ministry of Health

Assessment approach and implementation model

In 2020, the New Zealand (NZ) Ministry of Health, in collaboration with the World Economic Forum and Deloitte, developed the initial Digital Mental Health and Addiction Service (DMHAS) Evaluation Framework as a part of its response to the He Ara Oranga report – the government inquiry into mental health and addiction – and the urgent need for remote mental health services presented by the COVID-19 pandemic.

The DMHAS framework covers mobile apps and online tools and is driven by guiding principles and goals that reflect the needs of NZ, including Maori and Pasifika people.

The assessment approach is in the form of 55 wide-ranging individual standards that reflect the safety, quality, and specific needs of consumers. These standards are divided into mandatory questions, for all services, and supplementary questions, which are required depending on the functional scope, data collected, therapeutic claims and service workforce. Each standard contains an objective to meet, suggested evidence to show a vendor has met the standard, and examples of not meeting it.

The initially trialed implementation model is for the vendor to self-audit against the standards and submit their results and evidence to HealthNavigator, who act as the audit reviewer and keeper of a public accessible, pre-assessed library. A score is then calculated by linking each standard to an output of either safe, trusted, ethical, sustainable, and effective. Depending on the assessed level of the service, it may be fast-tracked..

Outcomes and lessons learned

The process was trialed in 2020 with several vendors and with HealthNavigator acting as the Auditor and Library owner.

- The preliminary questions created an effective advertising template to communicate to consumers.

- Vendors took action to address specific weaknesses such as a code of ethics, privacy and security, and focus on vulnerable communities.

- Vendors found guidance about what the market desired from digital mental health services to be very useful.

- Vendors found the self-audit process to be onerous and time-consuming to gather the evidence from multiple parties. Vendors suggested a "light" assessment for low-risk or early-stage services might be useful.

- The most useful outcome was to start conversations on governance for these services and use the results to identify topics for training and workshops to help vendors achieve improved outcomes such as: measuring clinical effectiveness, consumer-led co-design, data security, and promoting equity.

- Clarity and alignment of the consumer-facing scoring is extremely important to communicate to consumers and clinicians. "Usability" was not initially in the scoring and scoring terms must be clearly articulated.

The framework is currently undergoing assessment and public consultation, and is expected is expected to be released in June 2021.

Other governance and assurance model examples

Organization for Review of Care and Health Apps (ORCHA)[23] United Kingdom, Ireland, and the Netherlands	ORCHA is a private UK company that specializes in health app evaluation and advisory services. Its aim is to help governments, health and social care organizations to deliver health apps that will safely make the biggest impact in terms of improving outcomes. The specific review process is kept private. Revenue and funding are raised through app recommendation and matching, app evaluation and accreditation services, and development and integration advisory services.
National Safety and Quality in Digital Mental Health Standards (NSQDMH)[24] Australian Commission for Safety and Quality in Health Care (ACSQHC) Australia	This Australian public health entity is focused on contributing to better health outcomes and experiences. These standards were developed in 2020 through public consultation with a broad range of stakeholders and currently are published publicly and voluntarily only.
Evaluation and Assessment for Improving Access to Psychological Therapies Programme (IAPT)[25] National Institute for Health and Care Excellence (NICE) United Kingdom	NICE is public health body and agency of the NHS is charged with promoting clinical excellence in the NHS service. This program of evaluation is an in-depth, comprehensive, and peer-reviewed process to prepare specific solutions for inclusion in the NHS service.
American Psychiatric Association (APA) App Advisor[26] United States of America	The main professional organization of psychiatrists in the United States has published a publicly available model and set of evaluation questions to help clinicians unearth information to self-asses apps for inclusion in their practice.
OECD Recommendation on Responsible Innovation in Neurotechnology[27] Organization for Economic Co-operation and Development (OECD)	The OECD has published an aspirational and principles-based recommendation for responsible innovation in neurotechnology.
One Mind PsyberGuide[28] United States of America	This not-for-profit project aims to help people use technology to live a mentally healthier life. PsyberGuide has a publicly available assessment process based on credibility, transparency, user experience, and professional reviews, and a publicly viewable list of assessed apps.
e-compared European Union	The European project E-Compared conducts comparative effectiveness research in different mental care settings on the cost effectiveness of blended internet-based treatment for depression in comparison to standard care.
eticas European Union	eticas is a research and consulting organization that collaborates with other organizations to identify black box algorithmic vulnerabilities and retrains AI-powered technology with better source data and content.
Healthcare Information and Management Systems Society, Inc. (HIMSS) Unites States of America	HIMSS is a non-profit organization that acts as a global advisory and thought leader supporting the transformation of the health ecosystem through information and technology. HIMSS has initiatives to advance the safety and security of devices and systems through the Health Technology Alliance—with shared knowledge and idea exchange and public policy interpretation, advisory, and advocacy projects.

ImpleMentAll European Union	ImplementAll (IMA) is a European collaboration, funded by EU grants, toward faster and more effective implementation of eHealth interventions. IMA provides implementation support and research for mental health app assessment through frameworks such as the Normalization MeAsure Development questionnaire (NoMAD), based on Normalization Process Theory (NPT), and Model for Assessment of Telemedicine applications (MAST) -based mental health app assessment.
Node.Health United States of America	Node.Health is a not-for-profit organization in the United States that brings together a large network of stakeholders to advise, advocate for, research, and implement protocols for validation, clinical trials, publication, and integration for the digital innovation ecosystem.
Mobile App Rating Scale (MARS)[29] United States	The MARS is a multi-dimensional instrument for classifying and assessing the quality of mobile health apps. It focuses on engagement, functionality, aesthetics, and information quality. It can also be used to provide a checklist for the design and development of new high-quality health apps.

Contacts

Dr Stephanie Allen
Global Health Care Leader
Deloitte Global
steallen@deloitte.com.au

Dr Rohan Hammett
Asia Pacific Health Care Leader
Deloitte Australia
rhammett@deloitte.com.au

Marcus Schweizer
Manager
Deloitte Australia
mschweizer@deloitte.com.au

Endnotes

1.   Hickie, "The 'Uberisation' of Mental Health Care."

2.   Torous and Roberts, "Needed Innovation in Digital Health and Smartphone Applications for Mental Health."

3.   Torous et al., "Mental Health Mobile Phone App Usage, Concerns, and Benefits Among Psychiatric Outpatients."

4.   World Health Organization, "Mental Health Action Plan 2013-2020."

5.   World Health Organization.

6.   Trautmann, Rehm, and Wittchen, "The Economic Costs of Mental Disorders."

7.   Farmer and Stevenson, "Thriving at Work."

8.   Schwab and Vanham, "How to Prevent a Global COVID-19 Depression."

9.   Wan, "The Coronavirus Pandemic Is Pushing America into a Mental Health Crisis."

10. World Health Organization. "WHO | EHealth at WHO." World Health Organization. Accessed January 27, 2021. http://www.who.int/ehealth/about/en/.

11. World Economic Forum's Global Future Council on Mental Health, 2020

12. Fiske, Henningsen, and Buyx, "Your Robot Therapist Will See You Now."

13. World Economic Forum's Global Future Council on Mental Health, 2020

14. Naslund et al., "Digital Innovations for Global Mental Health."

15. Graham et al., "Artificial Intelligence for Mental Health and Mental Illnesses."

16. Fiske, Henningsen, and Buyx, "Your Robot Therapist Will See You Now."

17. Villongco and Khan, "'Sorry I Didn't Hear You.' The Ethics of Voice Computing and AI in High Risk Mental Health Populations."

18. Stiglic and Viner, "Effects of Screentime on the Health and Well-Being of Children and Adolescents."

19. Uusitalo, Tuominen, and Arstila, "Mapping out the Philosophical Questions of AI and Clinical Practice in Diagnosing and Treating Mental Disorders."

20. Fiske, Henningsen, and Buyx, "Your Robot Therapist Will See You Now."

21. Naslund and Aschbrenner, "Risks to Privacy With Use of Social Media."

22. U.S. Food & Drug Administration, "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device."

23. ORCHA, "Organization for Review of Care and Health Apps."

24. Australian Commission for Safety and Quality in Health Care (ACSQHC), "National Safety and Quality Digital Mental Health Standards."

25. National Institute for Health and Care Excellence, "Improving Access to Psychological Therapies (IAPT)."

26. American Psychiatric Association, "APA App Advisor."

27. Organization for Economic Co-operation and Development (OECD), "OECD Recommendation on Responsible Innovation in Neurotechnology."

28. One Mind, "One Mind PsyberGuide | The Mental Health App Guide Designed with You in Mind."

29. Stoyanov et al., "Mobile App Rating Scale."

Deloitte.