Canadian Centre on Substance Use and Addiction
Evidence. Engagement. Impact.

Centre canadien sur les dépendances et l'usage de substances
Données. Engagement. Résultats.

Mental Health Commission of Canada

Commission de la santé mentale du Canada

Balancing Promise and Risk:

Responsible AI in Mental Health and Substance Use Health Care



Dr. Kim Corace PhD CPsych
Chief Science and Innovation Officer
Canadian Centre on Substance Use and Addiction
Associate Professor, Psychiatry, University of Ottawa

Anique Turgeon, **Maureen Abbott**†
Mental Health Commission of Canada

Cesar Leos-Toro, Pam Kent, Jacqueline Fortier, **Kim Corace**†
Canadian Centre on Substance Use and Addiction

Learning objectives

Assess benefits and risks of AI in mental health and substance use health (MHSUH) care.

Describe the development of Canada's National Guidance for the safe, ethical and effective use of AI in MHSUH care

Apply the Guidance's conceptual framework to MHSUH care, responsible practice, and decision making.



Canadian Centre on Substance Use and Addiction
Evidence. Engagement. Impact.

Centre canadien sur les dépendances et l'usage de substances
Données. Engagement. Résultats.

Mental Health Commission of Canada

Commission de la santé mentale du Canada

Pressure on care

Growing need and limited capacity

Mental health and substance use health issues affect **1 in 5 Canadians** each year

Half of all Canadians will experience mental illness by age 40

Demand for services **exceeds system capacity**

Costs of substance use harms alone:
$49.1B CAD/ $35.3B USD

Significant Adoption

1 in 7 people living in Canada have used AI for mental health support

*Approximately **6 million** people living in Canada have used AI for mental health support.*

Use of AI tools for Mental health Support (past 12 Months)



Canadian Centre on Substance Use and Addiction
Evidence. Engagement. Impact.

Centre canadien sur les dépendances et l'usage de substances
Données. Engagement. Résultats.

Mental Health Commission of Canada

Commission de la santé mentale du Canada

G17. In the past 12 months, have you used any AI (artificial intelligence) tools to get advice or support for your mental health? Examples include: a chatbot that converses with you about your mood, a mood tracker that gives you automated feedback.

AI tools offer convenience and accessibility

People who use AI tools in Canada are drawn to them because they are low cost, available 24/7, and offer immediate access.

Motivations for using AI for mental health or emotional support

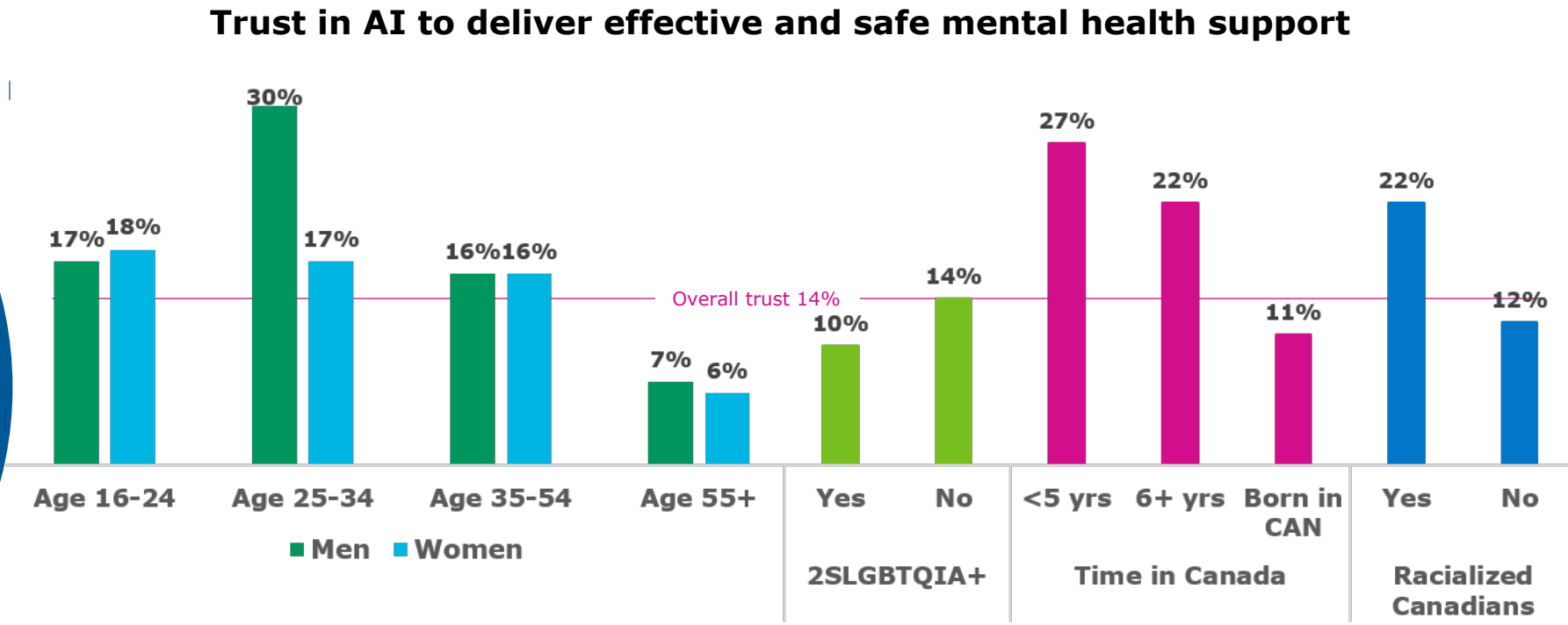


G21p27 What motivated you to use AI for mental health or emotional support? Select all that apply.

Only **14%** of people living in Canada **trust** AI for mental health support

*But trust is relatively **HIGHER AMONG EQUITY-DESERVING POPULATIONS** (young men, newcomers, and racialized groups living in Canada.)*

Limited Trust



G19. How much do you trust AI tools to provide helpful and safe mental health support?



Clinical care promises

Clinical capacity. Potential to reduce documentation burden and support more time to direct care.

Access and continuity. Potential to support monitoring and contact between appointments.

Care quality. Benefits depend on evidence, clinical fit, and integration with other services.

Need for clinical guidance

Supporting decisions in everyday practice



Clear expectations for providers



Practical supports for organizations



Care shaped by evidence and experience



AI across the care pathway

Different uses require different safeguards

Before care. Information, navigation, intake.

During care. Documentation and decision support.

Between visits. Monitoring and ongoing support.



AI is Here

Examples of Use

1. Chatbots/ Conversational AI agents
2. AI virtual therapists
3. Machine learning
4. Natural Language Processing Tools
5. Large language models
6. Deep learning models
7. Predictive modeling

Canadian Centre on Substance Use and Addiction
Evidence. Engagement. Impact.

Centre canadien sur les dépendances et l'usage de substances
Données. Engagement. Résultats.

Mental Health Commission of Canada

Commission de la santé mentale du Canada

Clinical risks and limitations

A plausible response may still be inaccurate or inappropriate

Accuracy. Errors, omissions, or unsupported recommendations.

Care decisions. Misalignment with assessment or delayed help-seeking.

Clinical context. Limited recognition of individual needs and circumstances.

Therapeutic relationships and shared decisions

AI use should support the person's voice and clinical judgement

Transparency. Explain where AI contributes to care.

Consent. Discuss limits, data use, and available alternatives.

Shared decisions. Patient centred preferences and goals.

Oversight. Preserve the ability to review, question, and override outputs

Equity in AI for MHSUH Clinical Services

Clinical benefit should be assessed across the people a service supports

Whose experiences?

Whose needs?

Whose voice?

DESIGNING WITH EQUITY IN MIND

***MHSUH:** Mental Health and Substance Use Health

Clinical benefits require evaluation

Assessment should include outcomes, experience, and full workflow

Use	Potential benefit	What a team should test
Documentation	Less time drafting notes	Errors, omissions, time spent checking
Intake	More timely access	Missed needs and differences across groups
Support between visits	Support while awaiting care	Unsafe advice and routes to human help
Decision support	Additional information for care	Fit with clinical assessment and client goals

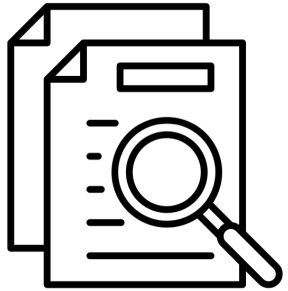
Organizational readiness for AI in clinical practice

Safe use depends on conditions surrounding the tool



CLINICAL FIT

Evidence for intended task and population



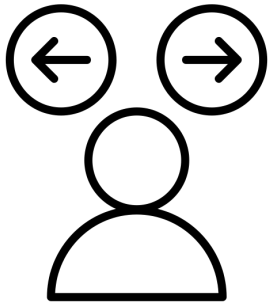
WORKFLOW

Training and time to review outputs



RESPONSE

Named responsibility and escalation routes



DATA & CHOICE

Approved data uses and accessible alternatives

Guidance for clinical implementation

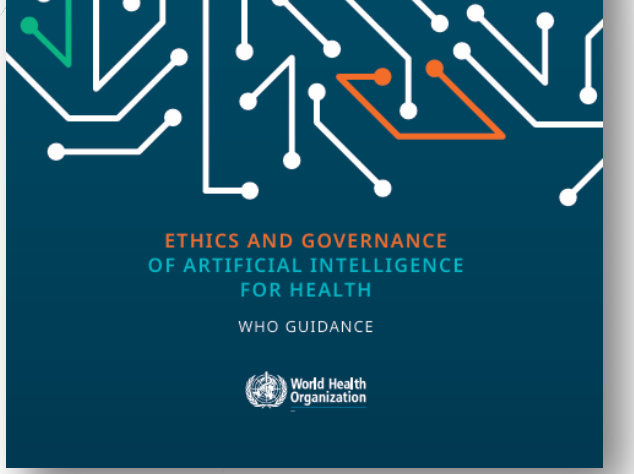
Connecting established principles to service-level decisions

Canadian Centre on Substance Use and Addiction
Evidence. Engagement. Impact.

Centre canadien sur les dépendances et l'usage de substances
Données. Engagement. Résultats.

Mental Health Commission of Canada

Commission de la santé mentale du Canada




Government of Canada | Gouvernement du Canada

Français

Search Canada.ca

MENU

Canada.ca › Departments and agencies › Health Canada › Drugs and health products › Medical devices

Transparency for machine learning-enabled medical devices: Guiding principles

On this page

- Background
- The guiding principles
- Contact us

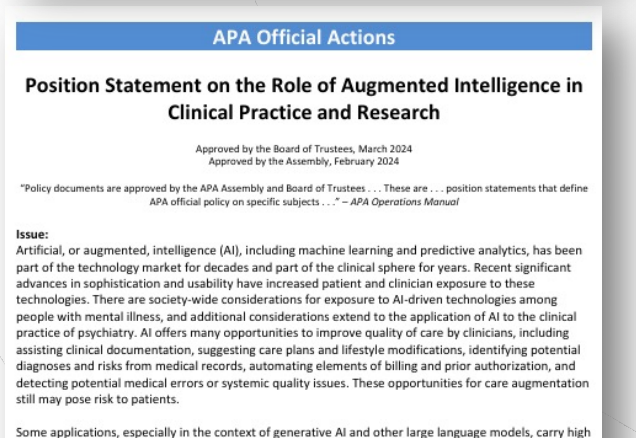
Background

In 2021, Health Canada, the U.S. Food and Drug Administration (FDA) and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA) jointly identified 10 guiding principles for good machine learning practice (GMLP). GMLP supports the development of safe, effective and high-quality artificial intelligence/machine learning technologies that can learn from real-world use and, in some cases, improve device performance.

Learn more:

- Good machine learning practice for medical device development: Guiding principles

Health Canada, FDA and MHRA have further identified guiding principles for transparency for machine learning-enabled medical devices (MLMDs). These principles build upon the GMLP principles, especially


APA Official Actions

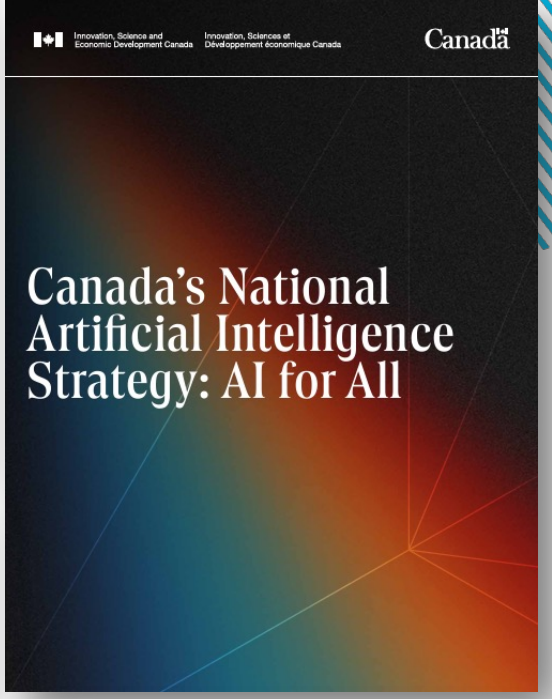
Position Statement on the Role of Augmented Intelligence in Clinical Practice and Research

Approved by the Board of Trustees, March 2024
Approved by the Assembly, February 2024

"Policy documents are approved by the APA Assembly and Board of Trustees . . . These are . . . position statements that define APA official policy on specific subjects . . ." – *APA Operations Manual*

Issue:
Artificial, or augmented, intelligence (AI), including machine learning and predictive analytics, has been part of the technology market for decades and part of the clinical sphere for years. Recent significant advances in sophistication and usability have increased patient and clinician exposure to these technologies. There are society-wide considerations for exposure to AI-driven technologies among people with mental illness, and additional considerations extend to the application of AI to the clinical practice of psychiatry. AI offers many opportunities to improve quality of care by clinicians, including assisting clinical documentation, suggesting care plans and lifestyle modifications, identifying potential diagnoses and risks from medical records, automating elements of billing and prior authorization, and detecting potential medical errors or systemic quality issues. These opportunities for care augmentation still may pose risk to patients.

Some applications, especially in the context of generative AI and other large language models, carry high




RESEARCH METHODS AND REPORTING

OPEN ACCESS

Check for updates

FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare

Karim Lekadir,[1,2] Alejandro F Frangi,[3,4] Antonio R Porras,[5] Ben Glocker,[6] Celia Cintas,[7] Curtis P Langlotz,[8] Eva Weicken,[9] Folkert W Asselbergs,[10,11] Fred Prior,[12] Gary S Collins,[13] Georgios Kaissis,[14] Gianna Tsakou,[15] Irène Buvat,[16] Jayashree Kalpathy-Cramer,[17] John Mongan,[18] Julia A Schnabel,[19] Kaisar Kushibar,[1] Katrine Riklund,[20] Kostas Marias,[21] Lameck M Amugongo,[22] Lauren A Fromont,[23] Lena Maier-Hein,[24] Leonor Cerdá-Alberich,[25] Luis Martí-Bonmatí,[26] M Jorge Cardoso,[27] Maciej Bobowicz,[28] Mahsa Shabani,[29] Manolis Tsiknakis,[21] Maria A Zuluaga,[30] Marie-Christine Fritzsche,[31] Marina Camacho,[1] Marius George Linguraru,[32] Markus Wenzel,[7] Marleen De Bruijne,[33] Martin G Tolsgaard,[34] Melanie Goisauf,[35] Mónica Cano Abadía,[35] Nikolaos Papanikolaou,[36] Noussair Lazrak,[1] Oriol Pujol,[1] Richard Osuala,[1] Sandy Napel,[37] Sara Colantonio,[38] Smriti Joshi,[1] Stefan Klein,[33] Susanna Aussó,[39] Wendy A Rogers,[40] Zohaib Salahuddin,[41] Martijn P A Starmans[33]; on behalf of the FUTURE-AI Consortium

For numbered affiliations see end of the article

Correspondence to: K Lekadir karim.lekadir@ub.edu (ORCID 0000-0002-9456-1612)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2025;388:e081554 http://dx.doi.org/10.1136/bmj-2024-081554

Accepted: 10 January 2025

Despite major advances in artificial intelligence (AI) research for healthcare, the deployment and adoption of AI technologies remain limited in clinical practice. This paper describes the FUTURE-AI framework, which provides guidance for the development and deployment of trustworthy AI tools in healthcare. The FUTURE-AI Consortium was founded in 2021 and comprises 117 interdisciplinary experts from 50 countries representing all continents, including AI scientists, clinical researchers, biomedical ethicists, and social scientists. Over a two year period, the FUTURE-AI guideline was established through consensus based on six guiding principles—fairness, universality, traceability, usability, robustness, and explainability. To operationalise trustworthy AI in healthcare, a set of 30 best practices were defined, addressing technical, clinical, socioethical, and legal dimensions. The recommendations cover the entire lifecycle of healthcare AI, from design, development, and validation to regulation, deployment, and monitoring.

SUMMARY POINTS

Despite major advances in medical artificial intelligence (AI) research, clinical adoption of emerging AI solutions remains challenging owing to limited trust and ethical concerns

The FUTURE-AI Consortium unites 117 experts from 50 countries to define international guidelines for trustworthy healthcare AI

The FUTURE-AI framework is structured around six guiding principles: fairness, universality, traceability, usability, robustness, and explainability

The guideline addresses the entire AI lifecycle, from design and development to validation and deployment, ensuring alignment with real world needs and ethical requirements

The framework includes 30 detailed recommendations for building trustworthy and deployable AI systems, emphasising multistakeholder collaboration

Continuous risk assessment and mitigation are fundamental, addressing biases, data variations, and evolving challenges during the AI lifecycle

FUTURE-AI is designed as a dynamic framework, which will evolve with technological advancements and stakeholder feedback

Introduction

In the field of healthcare, artificial intelligence (AI)—that is, algorithms with the ability to self-learn logic—and data interactions have been increasingly used to develop computer aided models, for example, disease diagnosis, prognosis, prediction of therapy response or survival, and patient stratification.[1] Despite major advances, the deployment and adoption of AI technologies remain limited in real world clinical practice. In recent years, concerns have been raised about the technical, clinical, ethical, and societal risks associated with healthcare AI.[2 3] In particular, existing research has shown that AI tools in healthcare can be prone to errors and patient harm, biases and increased health inequalities, lack of transparency and accountability, as well as data privacy and security breaches.[4-8]

To increase adoption in the real world, it is essential that AI tools are trusted and accepted by patients, clinicians, health organisations, and authorities. However, there is an absence of clear, widely accepted guidelines on how healthcare AI tools should be designed, developed, evaluated, and deployed to be trustworthy—that is, technically robust, clinically safe,

BMJ: first published as 10.1136/bmj.2024-081554 on 5 February 2025. Downloaded from https://www.bmj.com/ on 16 September 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Established Responsible AI Principles

- Safety
- Transparency
- Accountability
- Privacy
- Equity
- Human Oversight

The MHSUH Context

- Sensitive data
- Relational care
- Vulnerability
- Stigma
- Cultural context
- Complex clinical decisions

What does responsible AI look like in MHSUH practice?

Turning broad principles into practical guidance for real-world decisions

From Evidence to Guidance

2024–2025
Phase 1
- Planning

2025–2026
Phase 2
- Development of Conceptual AI Guidance

2026–2027
Phase 3
- Final AI Guidance
- Knowledge Translation

2027–2028
Phase 4
- Implementation Support
- Evaluation

Guidance to be released in 2027

Canadian Centre on Substance Use and Addiction
Evidence. Engagement. Impact.

Centre canadien sur les dépendances et l'usage de substances
Données. Engagement. Résultats.

Mental Health Commission of Canada

Commission de la santé mentale du Canada

Conceptual AI Guidance Framework



Trust & Explainability

- Understanding AI in practice
- Transparency
- Regular review & testing



Human-Centred Care

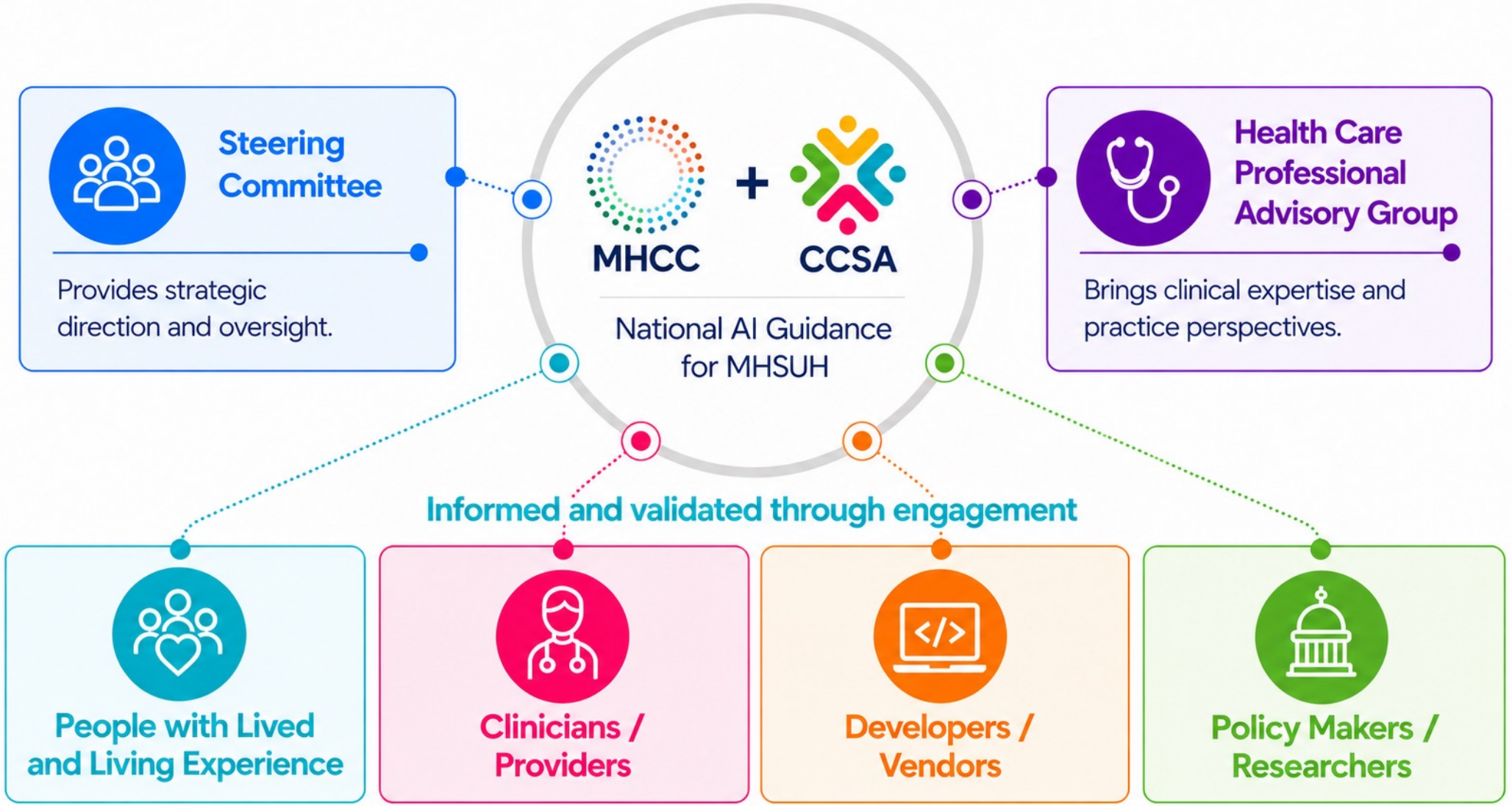
- Human oversight
- Informed consent
- Human autonomy & shared decision-making
- Personal data



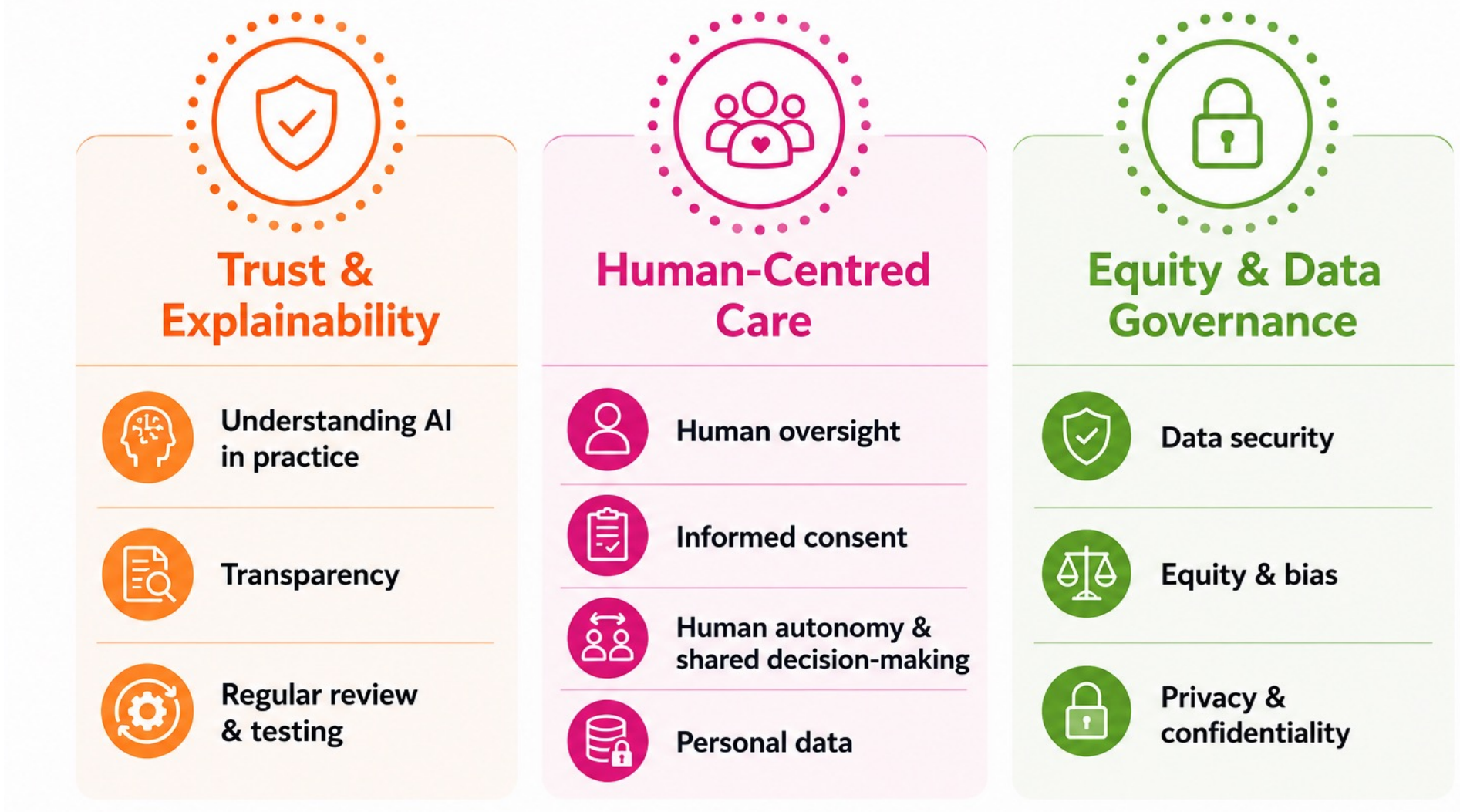
Equity & Data Governance

- Data security
- Equity & bias
- Privacy & confidentiality

Potential users help shape the rules



Three Critical Themes, Ten Considerations



Theme 1: Trust & Explainability

Understanding AI and assessing its performance

Consideration	For providers	For organizations
Understanding AI in practice	Know the task, limits, and common errors	Train staff for approved uses and updates
Transparency	Explain AI's role and limits in plain language	Require usable information from vendors
Regular review and testing	Check outputs and report concerns	Validate locally, monitor, and set stop rules

Theme 2: Human-Centred Care

Protecting relationships, choice, and clinical judgement

Consideration	For providers	For organizations
Human oversight	Review, override, and escalate concerns	Provide time, training, and authority
Informed consent	Check understanding and discuss choices	Make disclosure and alternatives workable
Human autonomy and shared decision-making	Keep the person's goals central	Protect access to human care
Personal data	Use approved systems and necessary data	Define access, retention, and reuse

Theme 3: Equity & Data Governance

Fair care and protection of sensitive information

Consideration	For providers	For organizations
Data security	Use approved tools and report concerns	Control access and prepare breach response
Equity and bias	Notice who is missed or misrepresented	Test across groups and involve service users
Privacy and confidentiality	Explain relevant data use and sharing	Limit access and prevent unapproved reuse

Translation to clinical decisions

Tests for implementation decisions



- ❑ Can staff explain its role?
- ❑ Do users understand its limits?
- ❑ What triggers retesting?

- ❑ Who reviews the output?
- ❑ Can people choose human care?
- ❑ Is consent meaningful?

- ❑ Who might be underserved?
- ❑ Which data uses are allowed?
- ❑ Who can challenge harm?

Equity in AI— who gets a say?



Choice in Care

People using services need accessible information and a human alternative



Influence on decisions

PWLLE have a role in choosing outcomes and reviewing harms



Authority over data

Relevant Nations and rights holders guide decisions about their data

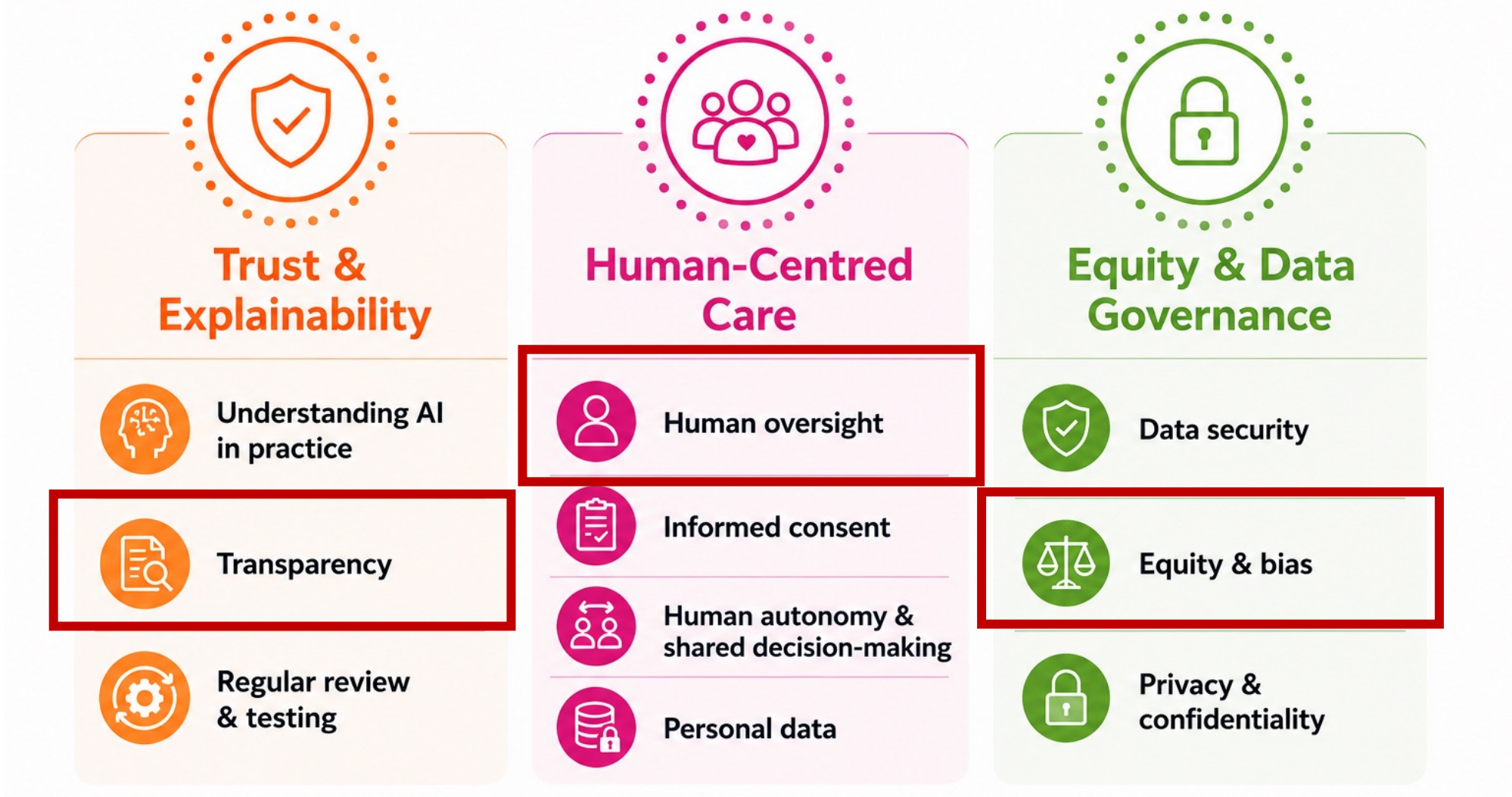


Agreements in practice

Organizations agree on data uses and community benefit with rights holders before adopting AI

First Nations Information Governance Centre. (n.d.). *The First Nations principles of OCAP®*. The First Nations Information Governance Centre.
Inuit Tapiriit Kanatami. (2018). *National Inuit strategy on research*. Inuit Tapiriit Kanatami.

Implications for clinical practice





Provider Responsibilities

Clinical judgement remains central to responsible AI use

Understand. Know the approved purpose and limitations.

Review. Check outputs before they influence care.

Communicate. Discuss AI use, choices, and concerns with patients.

Organizational Responsibilities

Making responsible use possible

Before. Define the use, assess evidence, and assign responsibility.

During. Support staff and monitor care quality, equity, and workload.

After changes or concerns. Investigate, retest, and revise or suspend use.



SHARED RESPONSIBILITY





Not a fixed rulebook

Jurisdictions can carry our method forward while adapting to local regulation, service realities, and communities



Responsible AI in clinical practice

Evidence-informed decisions. Human-centred care. Shared accountability.

What happens next?
Scan here to follow our journey.

Cesar Leos-Toro, Pam Kent, Jacqueline Fortier, **Kim Corace**[†]
Canadian Centre on Substance Use and Addiction

Anique Turgeon, **Maureen Abbott**[†]
Mental Health Commission of Canada