LAYING THE GROUNDWORK FOR

ARTIFICIAL INTELLIGENCE TO ADVANCE PUBLIC HEALTH IN CANADA

Prepared By:

Calvin Yip, MSc
Dalla Lana School of Public Health, University of Toronto

Laura C. Rosella, PhD
Dalla Lana School of Public Health, University of Toronto

Sarah Nayani, PhD
Dalla Lana School of Public Health, University of Toronto

Claudia Sanmartin, PhD
Statistics Canada

Deirdre Hennessy, PhD
Statistics Canada

Corresponding Author:
Laura C. Rosella (laura.rosella@utoronto.ca)

Report layout:
Hanako Smith, PhD

Statistics Canada Statistique Canada

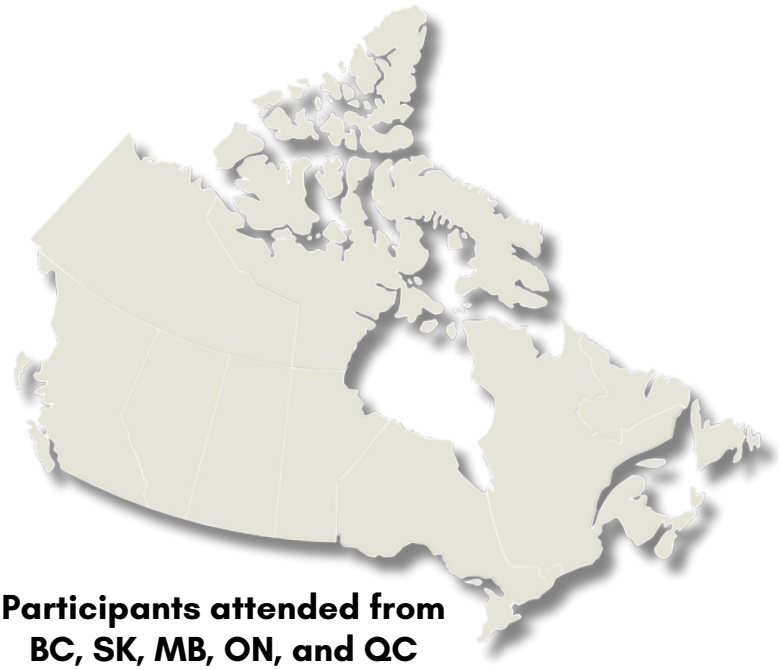
AI4PH AI FOR PUBLIC HEALTH

2024

TABLE OF CONTENTS

Executive Summary

Connection and understanding around national data, AI modelling, and public health



Participants attended from BC, SK, MB, ON, and QC

Artificial Intelligence (AI) has the potential to support public health in a wide range of applications, however, there are significant barriers to adopting AI in public health, and it is not yet clear which applications of AI will be most useful for public health research and practice. To elucidate and better understand barriers and uncertainties, as well as strengthen connections between national agencies and the academic community, a workshop entitled 'Connection and understanding around national data, AI modelling, and public health' was funded by the Canadian Institutes for Health Research (CIHR) and co-hosted by AI for Public Health (AI4PH) and Statistics Canada (StatCan) in Ottawa, Ontario on October 17th and 18th, 2023. The workshop connected public health stakeholders with experts in AI and public health, encouraging collaboration and facilitating discussion around the opportunities, priorities, and challenges associated with the use of AI for public health applications at the national level.

Workshop participants were individuals from national agencies including Health Canada, the Public Health Agency of Canada (PHAC), the Canadian Institute for Health Information (CIHI), and StatCan, public health scholars from across Canada leading research, training, and national data initiatives, and representatives from the AI4PH Community Advisory Board. AI4PH is a national training platform based at the University of Toronto's Dalla Lana School of Public Health and funded by CIHR. This training platform is building a workforce of public health researchers and practitioners who develop and apply innovations in equitable AI and machine learning to public health research, policy, and practice in Canada.

Executive Summary Continued

The main themes and recommendations for action that emerged from the workshop are:

Data and infrastructure: the need for improved data availability and access, common data standards, further linkages between data sources, and modernized analytical infrastructure

- **Recommendations:** address major infrastructure gaps related to linking health data for public health research and practice, with explicit consideration for data quality, data standards, and synthetic data; ensure that public health expertise and applications, distinct from clinical applications, are explicitly incorporated into national AI strategies and data initiatives

Skills and training: the need to address gaps in technical skills, the ability to work effectively in multidisciplinary teams, and dedicated strategies for training and upskilling

- **Recommendations:** develop a formal strategy to build AI expertise and upskill across the public health sector

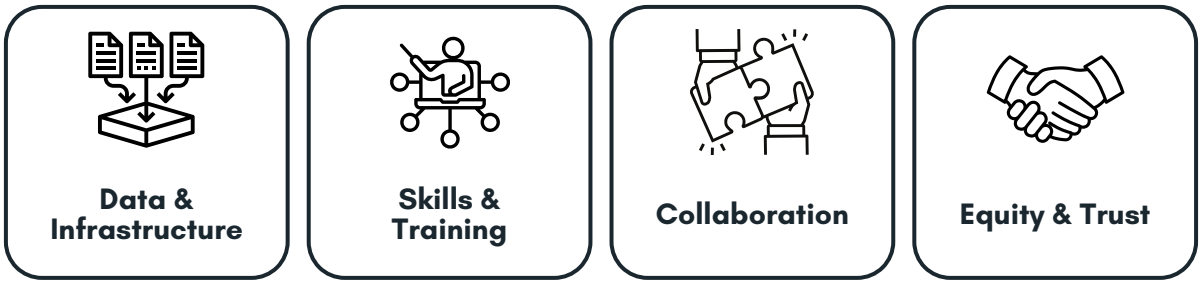
Collaboration: the need for more collaboration across disciplines and between organizations, and the importance of removing barriers to collaborative partnerships

- **Recommendations:** formally foster collaboration between academic and public health organizations to increase the likelihood that AI applications being developed are useful for public health; build mechanisms to foster collaboration and resource sharing across federal and provincial health and data agencies to increase efficiencies and accelerate learning

Equity and trust: the need for focused strategies to engage community members, build trust, mitigate bias, and improve health equity

- **Recommendations:** use formal frameworks and standards for AI adoption and oversight to ensure that AI applications in public health meet ethical, equity, and methodological standards

This report includes a more detailed synthesis of the main themes and recommendations to provide guidance for the responsible and effective adoption of AI in public health research and practice.



Introduction

Artificial Intelligence (AI) refers to a broad range of technologies that can perform tasks typically associated with the cognitive functions of humans (e.g., recognition, learning, logical reasoning).[1] Since the 1990s, AI research has largely focused on machine learning (ML) methods, which enable computer programs to learn from data and improve their ability to perform tasks. AI methods are well-established as powerful tools for data analytics in many sectors,[2] and are especially beneficial for new and emerging sources of complex data.[3] While potential applications of AI and ML in clinical medicine have been widely explored,[4] literature on the use of AI in public health remains very limited. Current applications of AI in health are largely limited to clinical settings designed for individuals. These applications often fail to capture social determinants of health, which are important for improving health equity through population level prevention programs or interventions.[5] With modern technological advancements, AI could potentially contribute to and improve public health strategies (e.g., surveillance, screening, risk prediction).[6,7] Examples of applications of AI in public health include the use of AI methods for detecting foodborne illness, estimating chronic disease risk, modeling the onset and trajectory of COVID-19-related outcomes, developing measurements for lifestyle risk factors, identifying patterns in healthcare service utilization, and causal inference.[8,9]

There are many risks, challenges, and limitations associated with using AI for public health, including the risk of exacerbating health inequities, challenges with model interpretability, structural challenges related to data sharing and analytical infrastructure, gaps in technical skills among the public health workforce, and other ethical and privacy concerns.[10] Consequently, there is a need to increase awareness of and strategically plan for the use of AI technologies to address population and public health challenges. There is a need for the agencies responsible for public health and national health data to come together and plan to address current barriers in collaboration with leading AI and public health scholars from across Canada.

The Centers for Disease Control and Prevention (CDC) describes 10 essential public health services that are needed to protect and promote the health of populations.[11] While it is acknowledged that some of these services (e.g., assessing the health status of a population, diagnosing health problems) could be supported by AI, it is not yet clear which applications will be most useful or how we could best scale these applications across public health settings. Additionally, significant barriers to adopting AI in public health research and programs have been identified. These include, but are not limited to, gaps in technical expertise, data availability and quality, and concerns of bias.[7, 12] In response to these uncertainties and barriers, the workshop was organized to address the opportunities, priorities, and challenges associated with using AI for public health applications at the national level.

The workshop was co-hosted by AI4PH and StatCan, and funded by CIHR. The overall aims of the workshop were to:

1. identify areas of connection between federal government departments, national organizations, academics, and public health stakeholders that can support the development and application of AI approaches for public health
2. identify the priorities of national stakeholders (i.e., government departments and national organizations) and the tasks they undertake that AI methods could support
3. learn from existing examples and experiences to identify facilitators and barriers to AI for public health, and increase knowledge and understanding of AI
4. identify specific training and capacity-building requirements to facilitate the use of AI at the national level
5. support new and strengthen existing connections between public health stakeholders and public health and AI scientists

Workshop Description

The workshop was held at StatCan's Simon Goldenberg Conference Centre in Ottawa, Ontario, on October 17th and 18th, 2023, and included a keynote presentation, two moderated panel discussions, three guided small group discussions (each focusing on one key area), and interspersed breaks throughout the days to encourage conversation and networking. The panel discussions focused on the opportunities, priorities, and challenges associated with the adoption of AI in public health; the first panel from an academic perspective and the second from a government perspective. The three areas for the small group discussions were data, people, and partnerships and engagement. Each small group had a facilitator to guide the discussion and a note-taker to summarize discussions.

As the first aim of the workshop was to identify areas of connection between federal government departments, national organizations, academics, and public health stakeholders, representatives from all of these groups were invited to attend. Of the 49 workshop attendees, most (84%) were situated in Ontario, while the remainder of attendees were from British Columbia, Saskatchewan, Manitoba, and Quebec. Most (59%) attendees were from government organizations, while non-government attendees were those from universities, research institutes, and the AI4PH Community Advisory Board (CAB) (Table 1). The CAB is comprised of members of the public with diverse lived experiences who have agreed to share their knowledge and input with the AI4PH leadership team. Members of the CAB are seen as equal partners in making recommendations and driving changes needed to achieve the goal of AI4PH, a national training platform based at the University of Toronto's Dalla Lana School of Public Health and funded by CIHR. This training platform was established to build a workforce of public health researchers and practitioners who develop and apply equitable AI and ML innovations to public health research, policy, and practice in Canada.[13]

Table 1. Primary Organizations of Workshop Attendees

Participants represented government, universities, hospital-based research institutes, and the community

Category	Organization
Government	Public Health Agency of Canada (12) Statistics Canada (12) Canadian Institute for Health Information (3) Health Canada (2)
University	University of Ottawa (4) McGill University (2) University of Toronto (2) University of Western Ontario (2) University of British Columbia (1) University of Manitoba (1)
Hospital-based research institute	Ottawa Hospital Research Institute (4) Trillium Health Partners (1)
General public	AI4PH Community Advisory Board (3)

Following the conclusion of the workshop, attendees were invited to provide feedback using an online evaluation form (Table S3 (Appendix B)). Of the 49 attendees, 25 (51%) participated in the workshop evaluation. Over two-thirds (68%) of respondents reported meeting at least 10 people for the first time, and all (100%) respondents reported that they were likely to contact at least one person from the workshop about a future collaboration (Figure S1 (Appendix C)). Most (⩾80%) respondents reported that they "agreed" or "strongly agreed" that the workshop increased their understanding of the tasks AI could support, facilitators and barriers to the use of AI, and the training and capacity-building requirements needed to facilitate the use of AI at the national level (Figure S2 (Appendix C)). All (100%) respondents reported that they "agreed" or "strongly agreed" that they made new connections or strengthened existing connections with peers working in their field.

Summary of Workshop Sessions

This section summarizes the structured sessions from the workshop. The full workshop meeting agenda is shown in Table S1 (Appendix A).

The state of public health in Canada – Identifying our major challenges (Keynote Presentation, Day 1)

Laura Rosella from the AI4PH Health Research Training Platform and the University of Toronto delivered the keynote presentation. Rosella offered context by highlighting ongoing public health challenges including widening socioeconomic disparities, climate change, and misinformation. She emphasized that unprecedented advancements in technology have opened new opportunities for applications of AI in public health, and discussed potential ways in which AI could help prevent disease, promote health, and reduce inequities through applications such as prediction, forecasting, and risk stratification. She concluded by recognizing that advancing the adoption of AI for public health requires collaboration between disciplines such as data science and public health, and between governmental and academic organizations.

Identifying opportunities, priorities, and challenges – Perspectives from academia (Panel Discussion, Day 1)

The first panel discussion was on the opportunities, priorities, and challenges associated with using AI for public health from an academic perspective. Dan Lizotte (AI4PH, Western University) moderated, and the panelists were David Buckeridge and Hiroshi Mamiya (AI4PH, McGill University), Lisa Lix (AI4PH, University of Manitoba), and Kim McGrail (AI4PH, University of British Columbia).

The priorities discussed related to the adoption of AI in academia and the importance of collaboration, and were detailed as:

- broader adoption and coverage of AI-related topics in university coursework in public health programs, including the incorporation of coding skills and foundational AI concepts into core curricula.
- broader adoption and expansion of AI-related topics and applications in public health, and support for this vision among the most senior leaders and administrators in academic institutions.
- ensuring that AI applications being researched and developed in academic settings are important and relevant to public health organizations, which requires collaboration across disciplines such as data science and public health, and between academic institutions and government organizations responsible for delivering public health services.

The opportunities discussed were related to expanding the adoption of AI-related teaching and research, leveraging modern technology, and fostering collaboration, and specified as:

- incorporating AI content into public health curricula in universities.
- developing and applying AI methods using the most up-to-date computing infrastructure, software tools, and computational resources.
- introducing incentives and structures that support collaboration, including internships and training platforms that provide opportunities for trainees to engage across organizations and help develop long-term partnerships that last beyond studentships.

The challenges discussed were related to teaching, upskilling, recruitment, and obtaining funding for work on AI, and included:

- incorporating flexibility into teaching to accommodate students from different disciplinary backgrounds (e.g., those with technical skills in statistics, mathematics, or computer science who also have subject matter knowledge in public health).
- maintaining up-to-date AI expertise in public health in the context of rapidly evolving AI technologies and methodologies.
- recruiting and retaining candidates with expertise and technical skills in AI.
- designating funding for AI work in the context of public health specifically, which differs from clinical and other applications of AI.

Identifying opportunities, priorities, and challenges – Perspectives from government (Panel Discussion, Day 1)

The second panel discussion was on the opportunities, priorities, and challenges associated with using AI for public health from a government perspective. David Buckeridge (AI4PH, McGill University) moderated, and the panelists were Chris Allison (PHAC), Kayle Hatt (Health Canada), Jeff Hatcher from (CIHI), and Claudia Sanmartin (StatCan).

The priorities identified related to improving data and addressing gaps in technical skills, and further described as the need for:

- improving data quality, consistency (e.g., through standardization), availability, linkage opportunities, and ease of data access.
- maintaining up-to-date technical skills by continuously reskilling and upskilling individuals in government organizations.

There was no consensus on which specific AI methods or applications are considered the most important, as methods of interest can vary greatly depending on the use case. Most applications of AI methods for public health in government remain in the developmental stage, and the results of these exploratory initiatives will help organizations develop their priorities. It was suggested that maintaining connections and sharing results and lessons learned among organizations is critical to advancing AI in the broader public health community.

The opportunities identified for the application of AI in government public health and data organizations were related to leveraging new data sources, applying new methods, and encouraging collaboration, and called for:

- generating synthetic data using AI to develop and test new methods without having to use real respondent or patient data.
- applying AI methods to existing and new linked data sources as they become available.
- connecting experts within the federal government to each other, research organizations, and policy makers at different levels of government (e.g., municipal, provincial, territorial) to encourage collaboration, resource sharing, and knowledge mobilization.

The challenges discussed were related to data, funding, ethics, and oversight, and included:

- data challenges and limitations related to data access, quality, and infrastructure, which hinder the advancement of work on AI.
- technical, logistical, and funding challenges related to upskilling employees in AI.
- a lack of dedicated funding for developing and implementing AI initiatives.
- ethical and equity concerns (e.g., data privacy, bias, inequity) associated with using AI for public health applications.
- the need for AI-specific frameworks and standards for oversight to help ensure that AI applications meet ethical standards (e.g., for data privacy) and advance health equity; government organizations should be ready to revisit and adopt formal regulations as they become available, and transparently describe how applications of AI are reviewed.

Alignment between academic and government perspectives

Several recurring themes were identified and discussed as important from academic and government perspectives, including:

- applying novel AI methods to gain additional insight from emerging high-dimensional data.
- upskilling to maintain up-to-date knowledge and skills related to the use of AI for public health applications, which is necessary to pursue innovative public health applications and will require dedicated resources.
- collaboration between organizations to increase the likelihood that AI applications being explored and developed are relevant and useful for public health.

In the context of AI for public health, the differences in perspectives between academia and government reflect key differences in their priorities and goals. Representatives from academia stressed the importance of expanding the adoption of AI-related topics in public health curricula to produce graduates with knowledge of both AI and public health. Representatives from government organizations focused heavily on the need to improve data quality, availability, and access to help advance work on AI for public health applications.

Format of the Guided Small Group Discussions

Small group discussions were employed to facilitate and strengthen connections and understanding among the national agencies and academics. In the fourth, fifth, and sixth sessions, attendees were organized into six groups to discuss areas as they relate to the application and uptake of AI in public health. The six groups were organized to evenly distribute attendees from academia and government across the groups. Each group included a facilitator and note taker, both of which were consistent for all sessions. Other group members were moved between groups for each session to allow attendees to meet and discuss with a greater number of people. Discussions were led by a series of guiding questions (Tables S2a, S2b, and S3c (Appendix A)). When reviewing summaries of these small group discussions, it should be acknowledged that the topics discussed were influenced to some extent by the earlier panel discussions and the guiding questions presented to attendees.

Data: Leveraging national and other data for advanced data science applications for health
(Guided Discussion on Data, Day 1)

In this session, attendees shared experiences and identified opportunities and challenges in using national level or other data for AI development and deployment across health settings. The main themes identified and discussed by the groups are summarized below.

Improved data availability and access

Concerns were raised about the length of time it takes to access data from Canadian government organizations (at both the federal and provincial level), noting that similar data in some other countries are available more quickly. In many situations, recent or real-time data are critical for identifying actionable insights. For example, public health events like the COVID-19 pandemic and the opioid crisis can evolve rapidly, and a lack of current data hinders Canada's ability to lead and innovate using newer AI methods for outcome detection, forecasting, or risk prediction. In academia, researchers face short timelines due to time-limited funding cycles, and often must prioritize research using currently available data. If timely health data are unavailable, researchers may apply their expertise in other areas or use less suitable data. Even when datasets become available, they can be difficult and time-consuming to access. Some attendees reported training models on US or UK data, which are often open source or much easier to access. In Canada, researchers must go through lengthy and complicated processes (e.g., proposal, justification, ethics, security clearance) to access administrative health data, while there are alternatives such as Medical Information Mart for Intensive Care (MIMIC) datasets with streamlined access processes,[14] which could serve as models.

Recommendations and potential next steps:

- explore ways to shorten the time between collecting data and making it available, acknowledging that this can be complex and resource intensive.
- improve data access by building on existing partnerships and data-sharing agreements with organizations that house population-level data (e.g., StatCan, CIHI, PHAC).
- generate synthetic data using AI methods to enable researchers to test methods and develop models without having to use real respondent or patient data.
- align efforts with the Pan-Canadian Health Data Strategy (pCHDS), which is focused on modernizing and streamlining health data collection, sharing, and access across the provinces and territories.[15]
- build on the pCHDS to consider the unique needs of research and AI-based applications.

Common data standards and further linkages between data sources

AI methods can analyze disparate data sources to uncover complex relationships, but a lack of standardization and linked data limits the use of these methods across datasets. For example, a specific variable may be defined or categorized differently across datasets. Even within the same data source, there are often changes in definitions, names, or categories of variables over time, which makes it difficult to compare findings across time. Attendees highlighted the value of linking data from different sources. For example, some attendees felt that rich survey data cannot reach its full potential without linkage with administrative health data or other valuable data such as environmental exposures or socioeconomic data. It was acknowledged that there are numerous barriers to data linkage, including cost, time, resources, governance, and permission. Additionally, identifying and finding linked datasets remains challenging, for example, one of the few known lists of linked government datasets is a list of datasets linked to the Derived Record Depository at StatCan.[16]

Recommendations and potential next steps:

- work towards common data standards for key variables of interest (e.g., geographical indicators, health outcomes) and consider connecting with Health Data Research Network (HDRN) Canada, an organization that is currently working towards multi-regional data to allow researchers to address health challenges that cross boundaries.[17]
- develop common guidelines for data linkage that consider re-identification risks to enable the linkage of more datasets.
- create more linked datasets (which may include linking additional datasets to existing linked datasets) and make them available to public health researchers.
- develop a publicly available list of linked government datasets for public health researchers.

Modernized analytical infrastructure

Several discussions reinforced the importance of investing in and modernizing analytical infrastructure. Attendees were interested in potential implementations of AI to speed up the data preparation process, and noted that AI could support or replace manual processes such as medical coding or manual data entry to help pre-process data or create datasets or metadata files. Additionally, AI methods can be used to search for and retrieve information from datasets and clean and prepare datasets for analysis. For example, the processing of unstructured text data such as social media posts and patient charts and notes can be sped up using natural language processing (NLP), which can automatically clean and standardize these data for quantitative analysis. It was stressed that the implementation of AI-based tools for data preparation would benefit greatly from machine-readable metadata. Upgrading computer hardware or investing in cloud infrastructure was also discussed. Participants provided examples of having insufficient computing resources (e.g., insufficient RAM (random-access memory)) to analyze very large datasets using complex methods. They also discussed having difficulty obtaining graphics processing units (GPUs) for graphics-intensive work like training models using images or audio. Even in the absence of compute limitations, IT operations and security may be rate-limiting factors for obtaining software, as approval and installation times for new software can be lengthy.

Recommendations and potential next steps:

- explore the use of AI and modern data engineering pipelines to extract data and speed up the data preparation process.
- ensure that metadata being created for new datasets are machine readable.
- advocate for and invest in compute infrastructure (e.g., physical hardware, cloud infrastructure) that meet the computing needs of those working on AI applications in public health.
- encourage communication between software users and IT administrators within and across organizations to ensure that there is mutual understanding of software needs and challenges.

People: Training and recruiting people and teams with appropriate technical, public health subject matter, or oversight skills (Guided Discussion on People, Day 2)

In this session, attendees shared experiences and identified opportunities and challenges in training and recruiting people and teams with appropriate technical, public health subject matter, and oversight skills. The main themes identified and discussed by the groups are described below.

Addressing gaps in technical skills and the ability to work effectively in multidisciplinary teams

Many attendees raised concerns about gaps in technical skills and the ability to work effectively in multidisciplinary teams on applications of AI in public health. One of the issues discussed was how difficult it is to find individuals with both advanced technical skills and public health subject matter knowledge. Technical expertise is needed to develop and implement AI applications, while public health expertise is needed to interpret results and provide context. A possible explanation for the observed gap in technical skills may be the lack of focus on relevant knowledge and skills in AI methods, open source programming, and data or computer engineering in universities and government. For example, some attendees noted that public health curricula in universities and current hiring processes in government organizations have focused on statistical software programs that differ from programs typically used for applications of AI. Additionally, recruiting individuals with advanced technical skills in research institutes and government organizations may be difficult because of the differences in compensation between the private and public sectors. Some noted a disconnect between developing complex technical AI content and translating findings to an applied public health or policy audience. This is a concern, as AI methods and policy issues are complex, and being able to tell a clear and compelling story is important for informing decision-making. The ability to effectively communicate across disciplines and integrate diverse perspectives is critical, as collaborations can include diverse team members with different priorities, perspectives, and risk tolerances.

Recommendations and potential next steps:

- incorporate the teaching of relevant knowledge and skills in AI concepts, methods, coding, and the critical appraisal of AI methods into public health curricula in universities.
- identify ways to encourage individuals to work in government organizations by, for example, communicating the non-monetary advantages of working in government organizations compared to the private sector.
- prioritize recruiting individuals with aptitudes like curiosity and motivation to learn in hiring processes if the desired technical skills are teachable.
- create a culture and expectation of continuous learning where employees maintain and build skills in conceptual knowledge, programming, and communication throughout their work life to ensure that skills are current.

A dedicated strategy for training and upskilling

Building capacity for training in AI was a recurring theme resulting from widespread concerns raised about the lack of AI knowledge and technical skills in the public health workforce. Financial resources and training providers were identified as key components required to build capacity for training. Even after securing funding to provide training in AI, it can be challenging to find training providers that meet the needs of public health professionals. For training to be effective, it must be tailored to the needs of employees and policy areas.

For example, mid-career employees may find the learning curve steeper than early-career employees who recently graduated from university programs that included training in the latest methods and software tools. It was recognized that upskilling strategies should be tailored to these different groups. The need for leaders' knowledge in this nascent area was also highlighted as being important. Directors, managers, and other leaders in public health need to know enough about AI concepts and developments to lead a team that works on AI, understand and interpret complex AI models, manage potential risks, and prioritize skills and areas to invest in.

Recommendations and potential next steps :

- ensure the availability of resources for ongoing training so that the skills of public health professionals can be updated throughout their careers.
- consider the specific needs of employees and policy areas when identifying and finding providers for training in AI.
- recognize the importance of upskilling directors, managers, and other leaders who may not directly apply AI methods but lead teams and set important strategic directions for employees working on AI projects.
- invite AI in public health experts to develop a competency framework to help guide academic institutes to incorporate key competencies into curricula (e.g., course material, fellowships, degree requirements, workshops).
- consider using the Core Competencies for Public Health in Canada, an example of an existing competency framework,[18] as a logical starting point from which to expand.

Partnerships and Engagement: Developing strategic partnerships and engagement within/between national agencies, between national agencies and academia, and with the community around AI for public health (Guided Discussion on Partnerships and Engagement, Day 2)

In this session, attendees shared experiences and identified opportunities and challenges in developing strategic partnerships and engagement within and between national agencies, national agencies and academia, and the community around AI for public health. The main themes identified and discussed by the groups are described below.

Increasing collaboration across disciplines and between organizations

There was unanimous agreement that collaboration between organizations is an important priority, while experiences with collaborations were mixed. The connections formed during collaborations are valuable for leveraging future projects. Attendees also supported the identification of gaps in skills within each organization, providing insight into specific skills gaps that should be addressed. Some attendees noted that the funding ecosystem for academia can be challenging, as funding agencies can be skeptical of AI methods they are unfamiliar with. More collaborations between researchers and government funding agencies can help address this concern by improving understanding of how AI methods can be used to address public health problems. Attendees repeatedly underlined the role of students as effective bridges between AI developments in academia and public health programs in government organizations through co-op, work-term, or internship opportunities. The importance of collaborating to create multidisciplinary teams was also highlighted. Machine learning operations (MLOps), for example, is an emerging field requiring expertise from several disciplines.[19]

Recommendations and potential next steps:

- increase awareness of how AI methods could be leveraged to improve public health research and practice across organizations and disciplines.
- continue using students and internships as vehicles for partnerships between organizations working on AI for public health.
- actively seek out collaborative partnerships with external organizations for multidisciplinary projects requiring expertise in both AI and public health.
- prioritize employee training opportunities through AI-focused organizations (e.g., Mila, Vector, AI4PH, Amii), potentially leading to future collaborations.

Removing barriers to collaborative partnerships

While collaborative partnerships are a priority in the AI and public health space, it was recognized that there are barriers to establishing these partnerships. The first barrier is a lack of awareness of what other groups (within and outside of an organization) are working on. One reason for this is a lack of communication between groups due to siloed work cultures, while another is that people from different fields may use different language to convey the same concepts. Consequently, groups may not be aware of the extent of overlap between their work and those from other disciplines. Lessons may be learned from the US Department of Health and Human Services, which maintains a central repository that lists all ongoing non-classified AI-related projects.[20] Another barrier is the complex process of establishing interorganizational collaborations. Collaborations across teams or organizations require negotiation, planning, and approval from each team, and additional hurdles related to rules for data sharing and finances often arise. For academic and government collaborations, it can take a long time for all government parties to approve of publications. Given the importance of publications in academia, this can have a negative impact on researchers. Lastly, when individuals and organizations have different roles, priorities, agendas, and incentives, the benefits of collaboration can conflict with the need or preference to protect IP, avoid risks associated with collaboration, or maintain a position or competitive advantage in a field.

Recommendations and potential next steps:

- increase awareness of what other groups (within and outside of an organization) are working on
- discourage the use of jargon and encourage working towards standardized terminology in discussions of AI and public health to improve common understanding across disciplines.
- where there are differences in priorities or agendas, encourage dialogue and discussion to reach common goals that would provide the greatest public health benefit to all Canadians.

A focused strategy for community engagement and building trust

While all agreed on the importance of community engagement and trust building, it was unclear how this was best achieved, and there was no consensus on what community engagement should look like. For example, it is not clear who in the community are the best representatives for broad public health issues or how many representatives are needed. Some attendees raised concerns that those who volunteer on community advisory boards may not represent the broader population, as they may be more likely to have the time and financial means to participate, which is a common challenge across patient and community engagement initiatives. Efforts must be taken to ensure that disadvantaged groups are represented. There were also concerns that community members may be recruited as token representatives who are not meaningfully engaged by the researcher. Attendees from academia highlighted concerns that meaningful community engagement requires significant time and resources and may not be valued as much as other academic outputs. Trust and transparency were raised as important considerations when describing AI methods. Sometimes related to IP protection, the description of AI-based models may not include enough detail about the methodology for the model to be reproduced or assessed by a third party. This makes it challenging to gain community trust in methods that are already widely seen as "black boxes".

Recommendations and potential next steps:

- encourage academic and government institutions to recognize researchers and programs for the importance and impact of their community engagement work.
- improve the community engagement process by reviewing and potentially adopting strategies from initiatives like the Strategy for Patient Oriented Research (SPOR) Evidence Alliance,[21] which has successfully integrated community engagement into its research processes.
- remove or reduce barriers to participating in community engagement activities when possible.
- strive for transparency in how AI methods are developed using approaches similar to the Government of Canada's Open Science initiative, which advocates for transparency through openly sharing data, tools, and research results.[22]

Dedicated approaches to mitigate bias and improve equity

Bias considerations are important in the development, assessment, and deployment of AI-based models, as biases in training data may be replicated in models. While the consensus was that the assessment of bias is important and necessary, attendees felt that assessing bias in AI-based models (and quantitative research more generally) is challenging. Concerns about systematic biases in data often focus on social determinants of health such as race and gender, and data on these variables are not always available. For example, it was noted that studies on bias in prediction models often focus on race and use US data, as the collection of data on race is not standardized in Canada. Even when data on variables such as race and gender are available, the sample size may not be sufficient to examine biases meaningfully. Insufficient sample sizes may lead to dichotomizing variables, resulting in the loss of information on heterogeneity within categories. Some attendees noted that it is sometimes necessary to leave biases in a model to understand why disparities exist, as models must reflect how real-world systems work. While this can provide insight to help understand and mitigate biases, individuals outside of epidemiology and public health may be skeptical about leaving biases in a model.

Potential recommendations and next steps:

- develop a responsible and comprehensive strategy for collecting information on key social determinants of health alongside health data, with guidance from initiatives such as StatCan's Disaggregated Data Strategy, which aims to continually identify and fill data gaps for variables such as race, immigrant status, sexuality, and gender identity in surveys.[23]
- ensure bias assessment is performed when developing AI applications, which includes describing biases observed in models or studies as transparently as possible when reporting results or recommendations (done in accordance with reporting guidelines where applicable) to provide knowledge users with information to consider these biases in their decisions, reducing the likelihood of adverse outcomes such as widening health inequities.
- clearly describe how bias was considered, assessed, and mitigated (if applicable) in applications of AI for public health.

Conclusions, Recommendations, and Next Steps

Leveraging AI for public health benefits is an important priority for national public health and health data agencies in Canada, public health practitioners, researchers, and the general public. This is a challenging effort that requires strategic planning and alignment between organizations and individuals aiming to make progress towards this goal. To lay the groundwork for the successful adoption of AI in public health, AI4PH and StatCan organized a workshop that connected national public health and health data agencies, AI and public health experts from academia, and community members. Having set out to establish 'Connection and understanding around national data, AI modelling, and public health', this CIHR-funded event succeeded in achieving these goals by connecting stakeholders and facilitating meaningful discussions. Participants had opportunities to discuss and analyze shared priorities and opportunities for the use of AI in public health, as well as the facilitators, barriers, and challenges associated with adopting AI for public health applications.

The major themes that emerged from these discussions were the need for: 1) improved data availability and access, 2) common data standards and further linkages between data sources, 3) modernized analytical infrastructure, 4) addressing gaps in technical skills and the ability to work effectively in multidisciplinary teams, 5) a dedicated strategy for training and upskilling, 6) increasing collaboration across disciplines and between organizations, 7) removing barriers to collaborative partnerships, 8) a focused strategy for community engagement and building trust, and 9) dedicated approaches to mitigate bias and improve equity. Importantly, these themes align closely with the priorities for the successful use of AI by public health organizations identified in the emerging literature.[10]

The workshop discussions yielded consensus recommendations to inform future action to enable the successful adoption of AI for public health applications. These recommendations are:

- address major infrastructure gaps related to linking health data for public health research and practice, with explicit consideration for data quality, data standards, and synthetic data.
- ensure public health expertise and applications, distinct from clinical applications, are explicitly incorporated into national AI strategies and data initiatives.
- develop a formal strategy to build AI expertise and upskill across the public health sector.
- formally foster collaboration between academic and public health organizations to increase the likelihood that AI applications being explored and developed are relevant and useful for public health policy and practice.
- build mechanisms to foster collaboration and resource sharing across federal and provincial health and data agencies aiming to advance AI for public health to increase efficiencies and accelerate learning across organizations.
- use formal frameworks or standards for AI adoption and oversight to ensure that AI applications in public health meet the highest ethical, equity, and methodological standards.

Next steps to address the themes highlighted in this workshop will include continuing discussions on AI in public health between the national agencies, the academic community, and the general public. These groups will explore ways to collaboratively address the challenges outlined in this report. AI4PH, in collaboration with PHAC and StatCan, is planning a dissemination event and panel discussion for this report in May 2024 (https://ai4ph-hrtp.ca/report-launch/). Some ongoing initiatives have also been identified as potential starting points for addressing some of the challenges described. Examples include the pCHDS for data governance, HDRN Canada for working towards common data standards, the SPOR Evidence Alliance for community engagement, and the StatCan Disaggregated Data Strategy for the collection of data to enable the consideration of bias and equity.

Acknowledgements

We would like to thank attendees for their time, enthusiasm, and contributions to the deep and insightful discussions throughout the workshop. We would also like to thank the speakers, moderators, note-takers, organizers, and all others who helped make this workshop possible. The workshop was funded by the CIHR Planning and Dissemination Grants – Institute of Community Support competition (Funding Reference Number: 189424) and in-kind support from StatCan who provided their conference facilities for this event.

Abbreviations

Table 2. List of abbreviations used in this report

Abbreviation	Unabbreviated term
AI	Artificial intelligence
AI4PH	Artificial Intelligence for Public Health
CAB	Community Advisory Board
CDC	Centers for Disease Control and Prevention
CIHI	Canadian Institute for Health Information
CIHR	Canadian Institutes of Health Research
GPU	Graphics Processing Unit
HDRN	Health Data Research Network
IP	Intellectual property
IT	Information technology
ML	Machine learning
MLOps	Machine learning operations
NLP	Natural language processing
pCHDS	Pan-Canadian Health Data Strategy
PHAC	Public Health Agency of Canada
RAM	Random Access Memory
SPOR	Strategy for Patient Oriented Research
StatCan	Statistics Canada
UK	United Kingdom
US	United States

References

1. Ma L, Sun B. Machine learning and AI in marketing–Connecting computing power to human insights. Int J Res Mark. 2020;37(3):481-504.
2. Rahmani AM, Azhir E, Ali S, et al. Artificial intelligence approaches and mechanisms for big data analytics: a systematic study. PeerJ Computer Science. 2021;7:e488.
3. Dobbins M, Buckeridge D. Artificial Intelligence in Public Health: Precision public health: Dream or reality? Canada Communicable Disease Report. 2020;46(6):160.
4. Deo RC. Machine Learning in Medicine. Circulation. 2015;132(20):1920-30.
5. Kassam A, Kassam N. Artificial intelligence in healthcare: A Canadian context. Healthcare Management Forum. 2020;33(1):5-9. doi:10.1177/0840470419874356
6. Benke K, Benke G. Artificial Intelligence and Big Data in Public Health. Int J Environ Res Public Health. 2018;15(12):2796.
7. Morgenstern JD, Rosella LC, Daley MJ, Goel V, Schünemann HJ, Piggott T. "AI's gonna have an impact on everything in society, so it has to have an impact on public health": a fundamental qualitative descriptive study of the implications of artificial intelligence for public health. BMC Public Health. 2021;21(1):40.
8. Sadilek A, Caty S, DiPrete L, et al. Machine-learned epidemiology: real-time detection of foodborne illness at scale. Npj Digit Med. 2018;1(1):1-7.
9. Artificial Intelligence for Public Health. Meet our trainees. Last revised in 2023. [Accessed April 11, 2024]. Available from: https://ai4ph-hrtp.ca/trainees/
10. Fisher S, Rosella LC. Priorities for successful use of artificial intelligence by public health organizations: a literature review. BMC Public Health. 2022;22(1):2146.
11. Centers for Disease Control and Prevention. 10 Essential Public Health Services. Last revised in 2023. [Accessed March 6, 2024]. Available from: https://www.cdc.gov/publichealthgateway/publichealthservices/essentialhealthservices.html
12. Corpuz JCG. Artificial intelligence (AI) and public health. J Public Health (Oxf). 2023;45(4):e783-e784.
13. Artificial Intelligence for Public Health. About AI4PH. Last revised in 2022. [Accessed January 19, 2024]. Available from: https://ai4ph-hrtp.ca/about/
14. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10(1):1.
15. Public Health Agency of Canada. Working with partners to modernize public health data. Last revised in 2024. [Accessed May 14, 2024]. Available from: https://www.canada.ca/en/public-health/programs/working-with-partners-modernize-public-health-data.html
16. Statistics Canada. Derived Record Depository (DRD) linkage status. Last revised in 2023. [Accessed January 16, 2024]. Available from: https://www.statcan.gc.ca/en/sdle/status
17. Health Data Research Network Canada. Strategic Plan 2021-2026. Published in 2022. [Accessed December 13, 2023]. Available from: https://www.hdrn.ca/wp-content/uploads/HDRN-Canada-Strategic-Plan_January-2022_comp.pdf
18. Public Health Agency of Canada. Core Competencies for Public Health in Canada. Last revised in 2023. [accessed December 13, 2023]. Available from: https://www.canada.ca/en/public-health/services/public-health-practice/skills-online/core-competencies-public-health-canada.html
19. Honkanen T, O'Dwyer J, Salminen V. Multidisciplinary Teamwork in Machine Learning Operations (MLOps). Human Factors, Business Management and Society. 2022;56:137-144.
20. US Department of Health and Human Services. Department of Health and Human Services: Artificial Intelligence Use Cases Inventory. Last revised in 2023. [Accessed December 13, 2023]. Available from: https://www.hhs.gov/about/agencies/asa/ocio/ai/use-cases/index.html
21. Gajic, A. When patients lead, research is relevant. Unity Health Toronto. Last revised in 2022. [Accessed January 17, 2024]. Available from: https://unityhealth.to/2022/09/patient-led-research/
22. Government of Canada. Open Science. Last revised in 2023. [Accessed January 17, 2024]. Available from: https://science.gc.ca/site/science/en/office-chief-science-advisor/open-science
23. Statistics Canada. Statistics Canada and disaggregated data. Last revised in 2023. [Accessed December 14, 2023]. Available from: https://www.statcan.gc.ca/en/trust/disaggregated-data

Appendix A - Workshop agenda and guiding questions

Table S1. Workshop agenda

Day 1 – TUESDAY, OCTOBER 17, 2023	
8:00 AM	**Breakfast**
8:30 AM	**Welcome and opening remarks** • Eric Rancourt and Claudia Sanmartin – Statistics Canada • Laura Rosella – AI4PH, University of Toronto
8:50 AM	**The state of public health in Canada - Identifying our major challenges** • Laura Rosella – AI4PH, University of Toronto
9:20 AM	**AI for Public Health: Identifying opportunities, priorities and challenges – perspectives from academia** • David Buckeridge – McGill University • Lisa Lix – University of Manitoba • Kim McGrail – University of British Columbia • Hiroshi Mamiya – McGill University
10:30 AM	**Networking coffee break**
11:00 AM	**AI for Public Health: Identifying opportunities, priorities and challenges – perspectives from government** • Chris Allison – Public Health Agency of Canada • Kayle Hatt – Health Canada • Kathleen Morris – Canadian Institute for Health Information • Claudia Sanmartin – Statistics Canada
12:15 PM	**Lunch**
1:00 PM	**Theme I - Data: Leveraging national and other data for AI and advanced data science applications for health** • Claudia Sanmartin – Statistics Canada
1:15 PM	**Break into small groups**
1:30 PM	**Guided Discussion on Theme I - in groups with AI4PH expert and note-taker**
2:40 PM	**Networking coffee break**
3:00 PM	**Reporting on individual group discussions, whole group discussion, reflection and questions**
Day 2 – WEDNESDAY, OCTOBER 18, 2023	
8:00 AM	**Breakfast**
8:30 AM	**Theme II - People:** Training and recruiting people and teams with appropriate technical, public health subject matter, or oversight skills • Doug Manuel – Ottawa Hospital Research Institute
8:45 AM	**Break into small groups**
9:00 AM	**Guided Discussion on Theme II - in groups with AI4PH expert and note-taker**
10:15 AM	**Networking coffee break**

10:45 AM	**Reporting on individual group discussions, whole group discussion, reflection and questions**
12:00 PM	**Lunch**
1:00 PM	**Theme III - Partnerships and Engagement:** Developing strategic partnerships and engagement within/between national agencies, between national agencies and academia, and with the community around AI for Public Health • Kim McGrail - University of British Columbia
1:15 PM	**Break into small groups**
1:30 PM	**Guided Discussion on Theme III -** In groups with AI4PH expert and note-taker
2:30 PM	**Networking coffee break**
2:45 PM	**Reporting on individual group discussions, whole group discussion, reflection and questions**
3:45 PM	**Wrap-up and next steps** • Laura Rosella - AI4PH, University of Toronto • Claudia Sanmartin- Statistics Canada • Laura Faye - Public health Agency of Canada

Table S2a. Guiding questions for the small group discussion sessions for Theme I

Theme I - Data: Leveraging national and other data for advanced data science applications for health

1. If you were/are interested in using national level or other Public Health data, how easy is it to find information or metadata about data sources?
2. If you have used national-level data or other Health data for model development:
    a. Was data access easier, e.g., open data, or more difficult, e.g., obtaining data required following a process?
    b. Was data easy to use once accessed, e.g., data standards, format, and structure were easy to understand/integrate into your analysis or model?
3. Are you aware of the data governance processes underlying access and use of data in national agencies or other agencies providing health data; could that data governance process be more transparent and easier to find information on?
4. If you are an academic, have you collaborated with national agencies (including but not limited to government departments, national data agencies, or charities) to use national-level data? If you are employed by a national agency, have you collaborated with academics or Public Health organizations to use data?
5. Wherever you accessed health data for model development, e.g., at your home organization or at a national agency, was modern analytical infrastructure available to you, e.g., sufficient computation powers and open-source libraries and programs?
6. From your perspective and in your role at a national agency, academic or other organization, can you identify priority areas or themes for data access, linkage across and within agencies, or development?
7. Can you share any examples of working collaboratively and successfully on national-level data?

Table S2b. Guiding questions for the small group discussion sessions for Theme II

Theme II – People: Training and recruiting people and teams with appropriate technical, Public Health subject matter skills or oversight skills.

1. From your perspective and in your role at a national agency or other organization, what are the biggest skill gaps (e.g., technical skills with data, analytics or modelling, subject matter skills, or oversight and team organizational skills)?
2. If your role includes hiring, what skills are hardest to find (e.g., technical skills with data, analytics or modelling, Public Health subject matter or oversight and team organizational skills)?
3. If your role includes hiring, what academic qualifications or skills best equip new employees (technical or otherwise) to fit easily into your organization? Related to this issue, if you are hiring in a national agency, would you be willing to provide academic institutions or training programs with a profile of an ideal candidate for this type of work?
4. If you are a technical, subject matter or oversight/organizational role at a national agency or other organization, what skills related to enabling AI development/deployment would you like to develop, is there opportunities within your organization to get the training you need?
5. Having collaborative projects between national agencies and other organizations is a potential way to foster knowledge exchange that benefits staff at both the national agency and in the academic or other organizations. In your experience, have collaborative projects furthered knowledge exchange and/or training goals, or have these collaborations been less fruitful than hoped?
6. How can the academic community support the training and skills development needed by national organizations? How can the agencies support continuing education and professional development?

Table S2c. Guiding questions for the small group discussion sessions for Theme III

Theme III – Partnerships and Engagement: Developing strategic partnerships and engagement within/between national agencies, between national agencies and academia and with the community around AI for Public Health.

1. From your perspective and in your role at a national agency, academic or other organization, what has been the greatest barrier to collaboration (e.g., lack of knowledge about what others in the AI4PH space are doing, assuming national organizations are not doing much, concerns about long processes for data access, concerns about lack of infrastructure for computation, others)?
2. Can you share any examples of successful partnerships, whether within/between national agencies, between national agencies and academia and with the community around AI for Public Health?
3. Beyond partnerships for the development and deployment of AI for Public Health applications, does your organization (e.g., national agency, academic or other organization) actively engage with members of the community to communicate how data is used in these applications and the intent and scope of the research?
4. Acknowledging that one of the guiding principles of Public Health practice is to improve health for all, does your organization (e.g., national agency, academic or other organization) have guidelines for research or model building that explicitly consider equity
5. Does your organization generate guidelines or make use of national or international guidelines for best practices in transparency and reproducibility for research and model building?

Appendix B - Workshop evaluation questions

Table S3. Questions included in the workshop evaluation form

Workshop Attendance
- Day 1 (in person)
- Day 2 (in person)

At the meeting how many people would you say you met for the first time?
- 0
- 1-5
- 5-10
- 10-15
- 15-20

Of the people you met, how many are you likely to contact after this meeting to collaborate with?
- 0
- 1-5
- 5-10
- 10-15
- 15-20

Of the people you met, please select the organisations that they represented (select all that apply).
- AI4PH
- AI4PH CAB Member
- Canadian Institute for Health Information
- Health Canada
- McGill University
- Public Health Agency of Canada
- Statistics Canada
- University of Ottawa
- University of Toronto
- University of Western Ontario
- Other (please specify)

At the meeting did you meet for the first time, anyone inside or outside your organization working on similar problems or similar areas to you?
- Yes
- No

At the workshop I increased my understanding of the tasks that AI methods could support.
- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree
- Don't Know/Not Applicable

At the workshop the examples shared helped me understand the facilitators and barriers to using AI in my practice.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree
- Don't Know/Not Applicable

At the workshop I increased my understanding of the training and capacity-building requirements needed to facilitate the use of AI at the national level.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree
- Don't Know/Not Applicable

At the workshop I made new and/or strengthened existing connections with my peers working in my field.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree
- Don't Know/Not Applicable

Please comment on the event overall and aspects both positive or negative that you would like to highlight (open-ended).

Appendix C – Workshop evaluation results

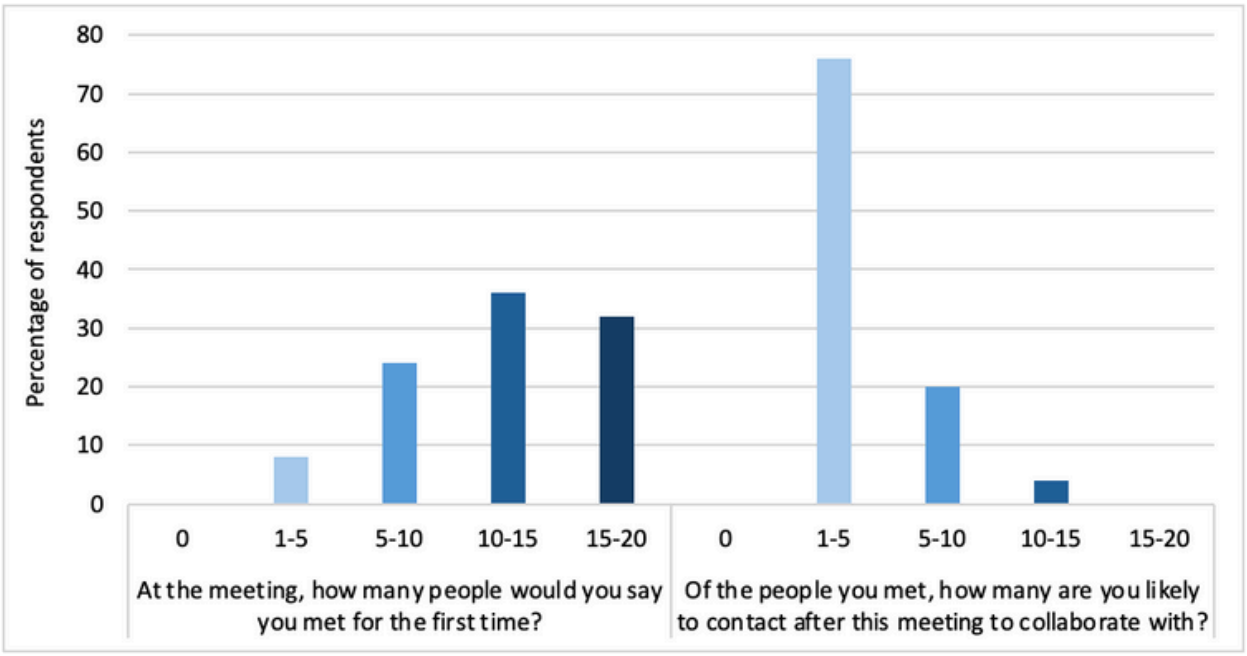


Figure S1. Number of people respondents reported meeting for the first time and number of people respondents reported being likely to contact after the workshop to collaborate with

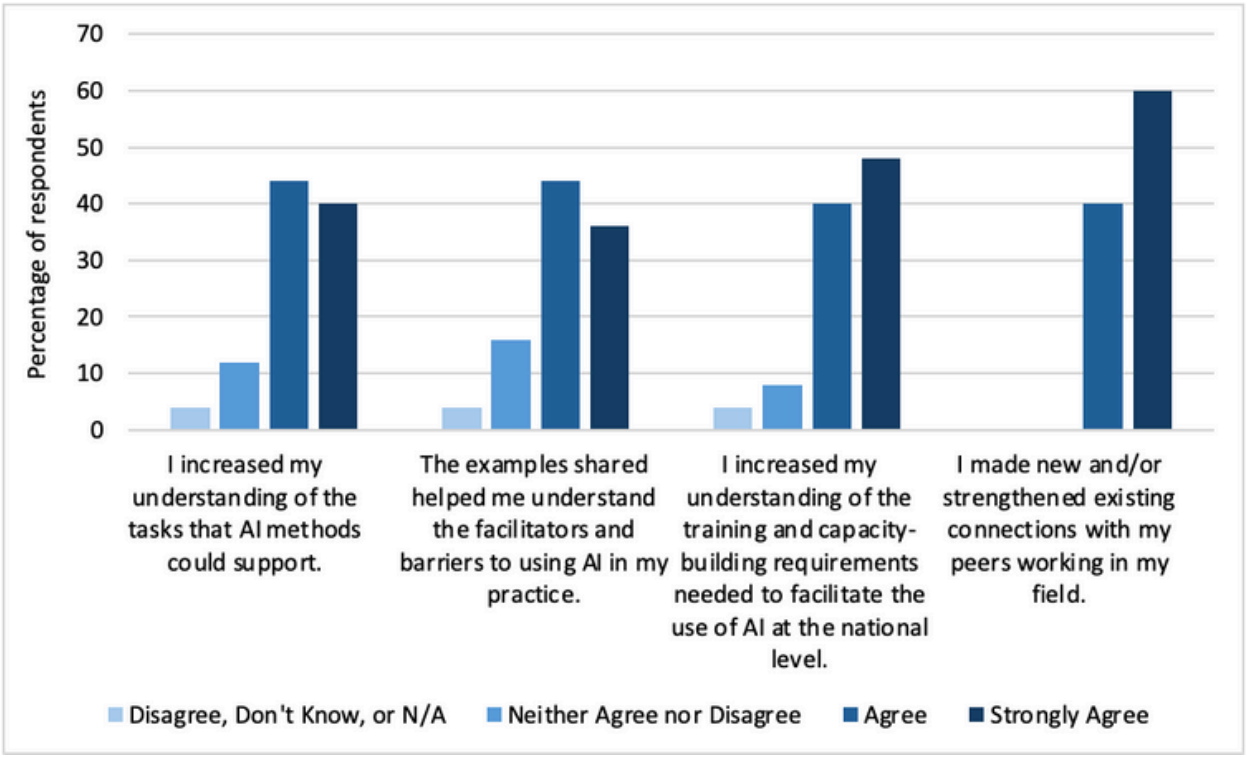


Figure S2. Respondents' opinions on how the workshop affected their understanding of AI in public health and their connections with peers working in the field

AI4PH AI FOR PUBLIC HEALTH

More Information About Us

AI4PH is focused on building capacity in AI and big data skills for transformative change in addressing population and public health challenges, and understanding how these tools impact health equity.

Our mission is to establish a workforce of public health researchers and practitioners who develop and apply innovations in equitable artificial intelligence (AI) and machine learning (ML) to public health research, policy, and practice in Canada.

For more on our traineeship programs, public courses, and network activity please visit:

www.ai4ph-hrtp.ca

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada

AI4PH is supported by the Canadian Institutes of Health Research (CIHR).

Contact Us :

Email Address
ai4ph.dlsph@utoronto.ca

Website
www.ai4ph-hrtp.ca

LinkedIn
/ai4ph-health-research-training-platform

Twitter
@ai4ph