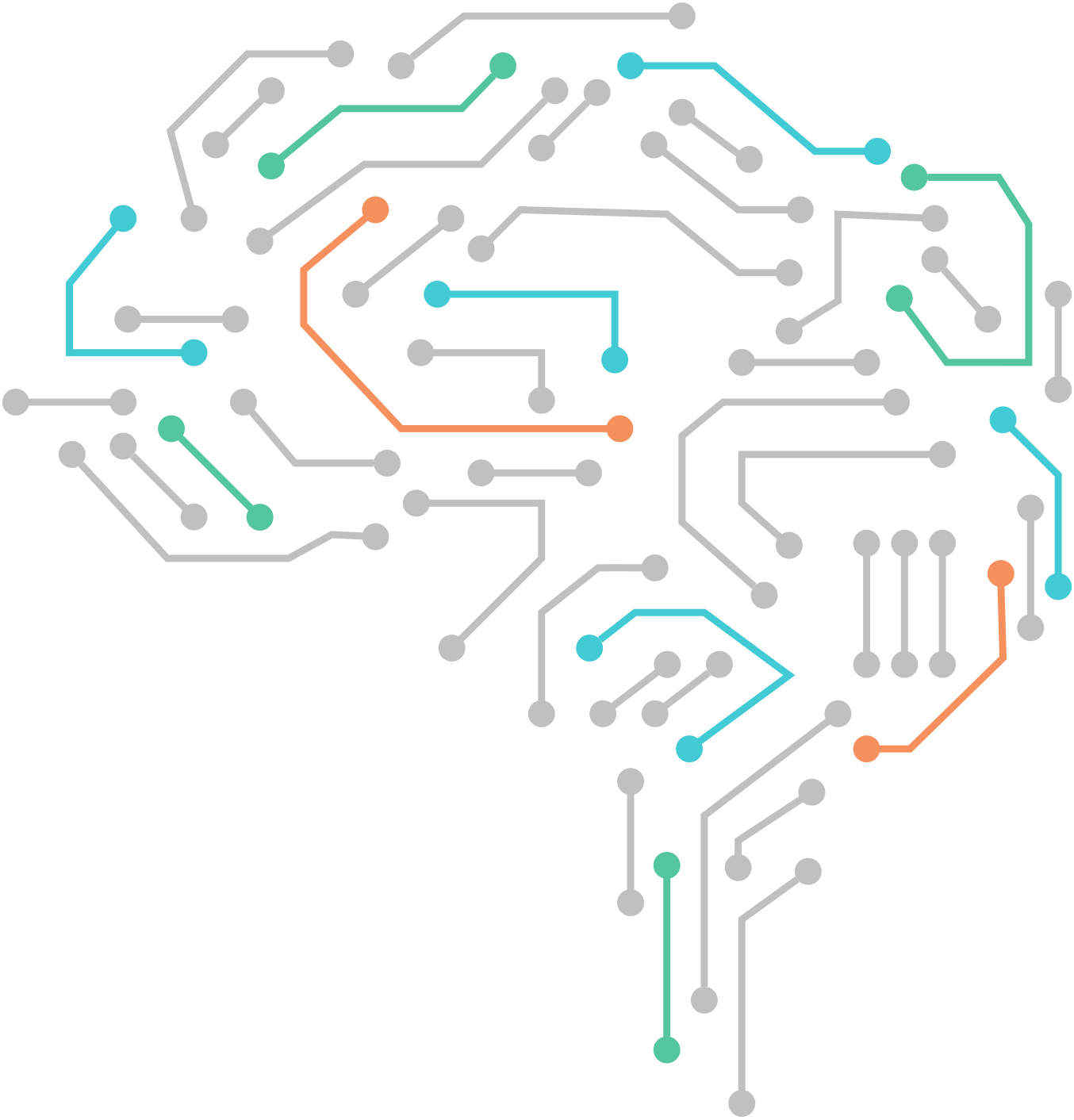
GENERATING EVIDENCE FOR ARTIFICIAL INTELLIGENCE-BASED MEDICAL DEVICES: A FRAMEWORK FOR TRAINING, VALIDATION AND EVALUATION

World Health Organization



GENERATING EVIDENCE FOR ARTIFICIAL INTELLIGENCE-BASED MEDICAL DEVICES: A FRAMEWORK FOR TRAINING, VALIDATION AND EVALUATION

World Health Organization

Generating evidence for artificial intelligence-based medical devices: a framework for training, validation and evaluation

ISBN 978-92-4-003846-2 (electronic version)
ISBN 978-92-4-003847-9 (print version)

Suggested citation. Generating evidence for artificial intelligence-based medical devices: a framework for training, validation and evaluation. Geneva: World Health Organization;2021. Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

Design and layout by Inis Communication

CONTENTS

List of figures

List of tables

FOREWORD

Artificial intelligence (AI) has potential to optimize the delivery of healthcare and improve outcomes for all. For countries which have yet to achieve universal health coverage, data-driven technology will play a vital role in the next decade. Current AI, machine learning and deep learning applications include the use of clinical decision support tools, diagnostics, and workflow optimisation solutions. AI is also being used to enhance health research and drug development, and in assisting with the deployment of different public health interventions, such as disease surveillance, outbreak response, and health systems management.

AI could greatly benefit low- and middle-income countries, especially in those countries that may have significant gaps in health care delivery and services. AI-based tools and data-driven technology as a whole could help governments extend health care services to underserved populations, improve public health surveillance, and enable healthcare providers to better attend to patients and engage in complex care.

For AI to have a beneficial impact on public health and medicine, ethical considerations must be placed at the centre of the design, development, and deployment of AI technologies for health. The evidence generated from the development and deployment of these devices must be robust and transparent, supporting claims for safety and performance. AI must be generalisable and work to improve outcomes for all populations. Existing biases in healthcare based on race, ethnicity, age, socioeconomic status and gender, that are encoded in data used to train algorithms, must be overcome.

Those same standards for development, deployment and post-market surveillance of AI tools must be applied in the global health context, especially in LMIC populations where governance and regulatory structures for the use of these devices is still evolving. This framework serves as a foundation document and considers minimum requirements for clinical evidence generation in three phases: 1) Software Development, 2) Software Validation and Reporting, and 3) Deployment and Post-Market Surveillance. It uses cervical cancer screening as a use-case to demonstrate the evidence generation considerations. This use-case is appropriate, given the enormous task ahead to eliminate cervical cancer, which remains one of the most common cancers and causes of cancer-related death in women across the globe, even though It is a preventable disease.

As recognised in WHO's *Global strategy to accelerate the elimination of cervical cancer as a public health problem,*