



Rwanda

CENTRE FOR THE FOURTH INDUSTRIAL REVOLUTION

A Partner of the World Economic Forum Network
for Global Technology Governance



Chatbots RESET Framework: Rwanda Artificial Intelligence (AI) Triage Pilot

INSIGHT REPORT
MARCH 2022



Contents

Foreword	3
Executive summary	4
Introduction	5
1 Chatbots RESET framework	6
1.1 Principles	6
1.2 Actions	7
2 Rwanda pilot: purpose and process	8
2.1 Purpose of the pilot	9
2.2 Piloting process	9
3 Rwanda pilot: lessons	10
3.1 Why did you decide to pilot?	10
3.2 Which healthcare application did you pilot?	11
3.3 Which AI technology did you use?	12
3.4 How was this AI technology developed for the pilot?	13
3.5 What are the challenges and opportunities in your use of AI in this application and in Rwanda?	13
3.6 What was your approach to governance in this pilot?	14
3.7 What is your governance process?	17
3.8 Which principles and actions from the Chatbots RESET framework did you choose? Why did you choose them?	17
3.9 How did you implement the actions?	18
3.10 What is your feedback on the Chatbots RESET framework?	21
3.11 Would you like to share any best practices?	21
Conclusion	22
Contributors	23
Endnotes	24

Disclaimer

This document is published by the World Economic Forum as a contribution to a project, insight area or interaction. The findings, interpretations and conclusions expressed herein are a result of a collaborative process facilitated and endorsed by the World Economic Forum but whose results do not necessarily represent the views of the World Economic Forum, nor the entirety of its Members, Partners or other stakeholders.

© 2022 World Economic Forum. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

Foreword



Crystal Rugege
Managing Director, Centre
for the Fourth Industrial
Revolution Rwanda

The COVID-19 pandemic triggered the rapid and large-scale adoption of artificial intelligence in healthcare, as every health system in the world underwent the ultimate stress test. While many healthcare providers had been reluctant to use chatbots in the past, the pressure of the pandemic demanded the deployment of new technologies in overwhelmed and under-resourced environments. Although healthcare systems in some countries have returned to normal, this high-stress scenario is still business as usual for many others around the world. Chatbots have the potential to close critical gaps in healthcare delivery and improve patient outcomes, particularly for underserved communities with resource limitations. However, we must acknowledge the risks that accompany the adoption of this technology in a healthcare setting and the disproportionate impact if something goes wrong.

Rwanda has been a pioneer in the strategic use of technology to increase access to healthcare.

We welcomed the opportunity to pilot the Chatbots RESET framework as the country launched its first AI-enabled triage service in partnership with Babyl. This pilot plays a vital role in ensuring the technology is deployed responsibly. In particular, one objective of this pilot was to expand the perspective of the project community by applying a Rwanda-specific lens to the framework, where the interpretation of principles such as safety, transparency and accountability need to be grounded in the local context. Indeed, this pilot is of notable significance since the AI-enabled triage service will eventually be accessible to most of the population through community-based health insurance, demonstrating the utility of chatbots in high-tech and low-tech settings.

There is still much to be learned about the benefits and risks of chatbots in healthcare. This framework is an important contribution to the global discourse. We hope this pilot has brought us a step closer to responsibly increasing access to healthcare.

Executive summary

Rwanda is leading the use of artificial intelligence responsibly to improve healthcare access for its citizens using a framework developed by the World Economic Forum.

Artificial intelligence (AI) is being adopted in healthcare for applications including precision medicine, standardizing information, intelligent personal health records, robotics and ambient assisted living. Automation and the speed of handling data and information enabled by AI in these applications have opened up opportunities to serve more patients. Symptom checkers are salient examples of the use of AI to triage patients rapidly.

With the increasing use of AI in healthcare and the lag in regulations needed to govern its use, alternative mechanisms are needed. In cooperation with global experts, the World Economic Forum has created a “soft governance” framework to facilitate the responsible use of AI (chatbots, in this case) in healthcare. The Chatbots RESET framework contains AI and healthcare principles to follow to govern the responsible use of chatbots in healthcare. It also includes actions that technology developers, medical providers and government regulators can take to implement the principles.

Two healthcare start-ups, a large hospital system and a global technology provider piloted the Chatbots RESET framework in 2021. This report summarizes a fifth pilot conducted in Rwanda in an AI triage service application. The technology developer and provider of the triage service, Babylon, chose three principles from the framework (safety, accountability and transparency) and implemented the associated actions with the principles.

This report summarizes the pilot in the form of a series of questions and answers with the service provider. The experience and lessons learned from this pilot and the preceding four pilots reinforce the benefits of using the Chatbots RESET framework to govern the responsible use of AI in healthcare. For private and public sector organizations concerned about the lack of concrete regulations for the use of AI in healthcare applications, the pilot outcomes point the way forward for the adoption of AI.

Image credit: Babylon



Introduction

Chatbots RESET is a multistakeholder framework for the responsible use of chatbots in healthcare.

Artificial intelligence (AI), among other Fourth Industrial Revolution technologies, has seen rapid adoption rates, with governments lagging in regulating the technologies to protect society from potential negative consequences of their use in areas of high public sensitivity, such as healthcare. There is a strong need for alternative “soft regulation” mechanisms to bridge the ensuing governance gap. The World Economic Forum has taken a systematic approach to developing such soft regulation mechanisms in the form of frameworks and toolkits for the responsible use of AI in several applications, including procurement, human resources and the use of facial recognition. Chatbots RESET is one such framework.

Working with a multistakeholder project community, the World Economic Forum has developed and published Chatbots RESET, a framework for the

responsible use of chatbots in healthcare. The development of the framework is in response to the increasing adoption of AI – especially chatbots – in healthcare settings worldwide. While the use of chatbots in various healthcare situations was being experimented with before, the COVID-19 pandemic created urgency in disseminating information quickly to the public. Chatbots were widely deployed for this purpose.

The framework contains a set of AI and healthcare principles and a set of actions for each of these principles. The actions provide recommendations for different players in the system using chatbots in healthcare: developers (who create the technology), providers (who use them in healthcare settings) and regulators (who are responsible for protecting the public).

Piloting

An important way to quickly validate the Chatbots RESET framework is to pilot the framework in real-life applications. Each organization piloting the framework selected a subset of the framework principles and tested them in their application context. Working with four partners, the Forum piloted the framework from January to April 2021 and published the results¹ in November 2021. Since then,

the Forum has been working with global healthcare company Babylon and its Rwandan entity Babyl to implement key principles from the framework for an AI triage tool. This tool is part of Babyl’s partnership with the Government of Rwanda to provide an integrated digital-first health service that leverages the country’s universal health insurance scheme, giving millions of patients access to healthcare in Rwanda.



Ethical use of AI is our responsibility – most of all in areas like healthcare.

Kay Firth-Butterfield, Head of Artificial Intelligence and Machine Learning,
World Economic Forum

1

Chatbots RESET framework

The framework consists of a set of principles and actions to operationalize each principle.

Following a landscape review of the use of chatbots in healthcare, the World Economic Forum convened an expert community from various parts of the system using these chatbots. Working closely with

community members, the framework was co-created during 2020. Below is a summary of the framework, the full details of which can be found in the original publication.²

1.1 Principles

The project community selected 10 principles for the Chatbots RESET framework from AI and healthcare principles based on their relevance and applicability and specifically interpreted them in the context of the use of chatbots in healthcare.

FIGURE 1 Examples of principles and their interpretations in the context of the use of chatbots in healthcare



Safety/non-maleficence

- The actions of chatbots shall not result in avoidable harm to humans or other unintended consequences, including deception, addiction or lack of respect for diversity



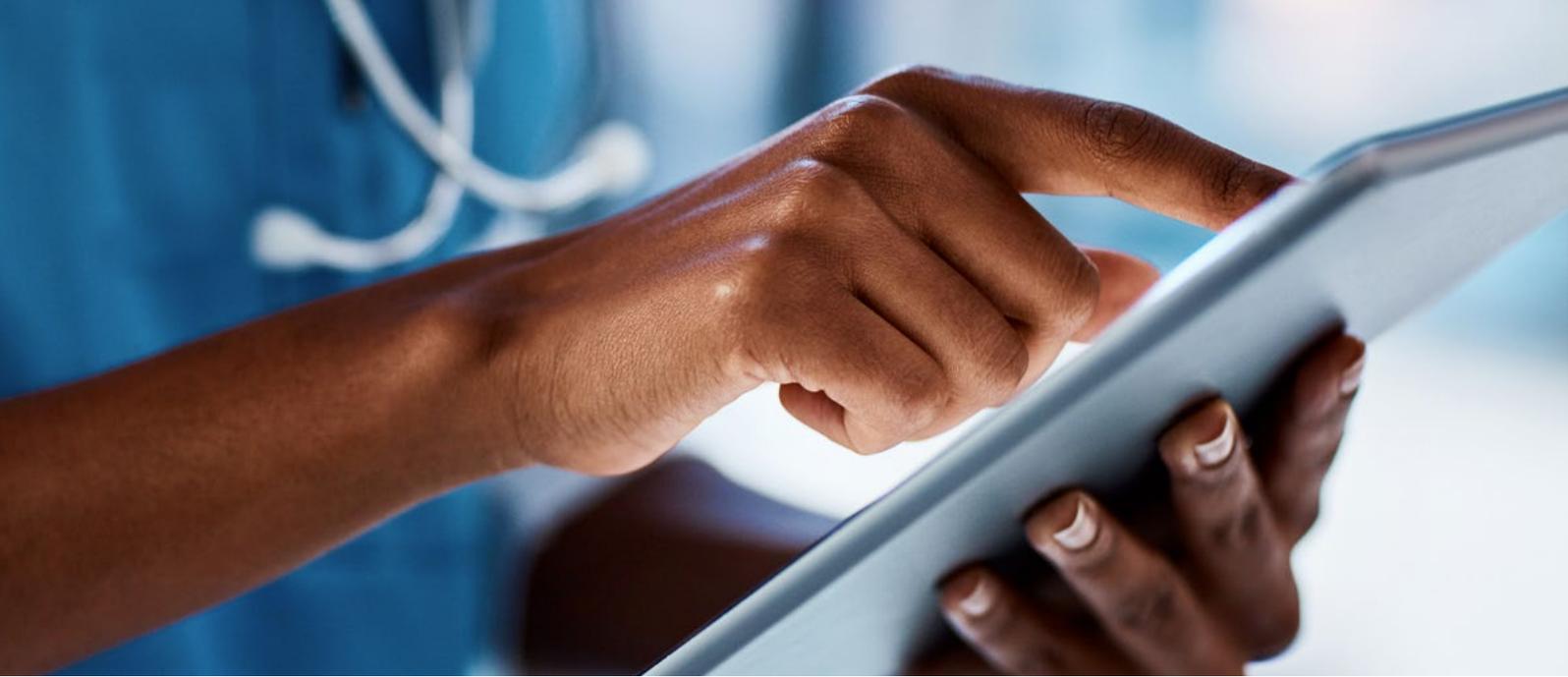
Accountability

- An entity (person or group) in the organization shall be accountable for the governance of chatbots
- Conclusions and recommendations of chatbots shall be auditable



Transparency

- Chatbot users shall at all times be made aware of whether they are interacting with an AI or a human or a combination of the two
- Chatbots shall clearly inform users about the limits of the system's performance, except in situations where not informing is required for the intended purpose of the chatbot
- Chatbot users shall be immediately informed if the chatbot is unable to understand the user or is unable to respond with certainty, except in situations where such communication interferes with the intended purpose of the chatbot



1.2 Actions

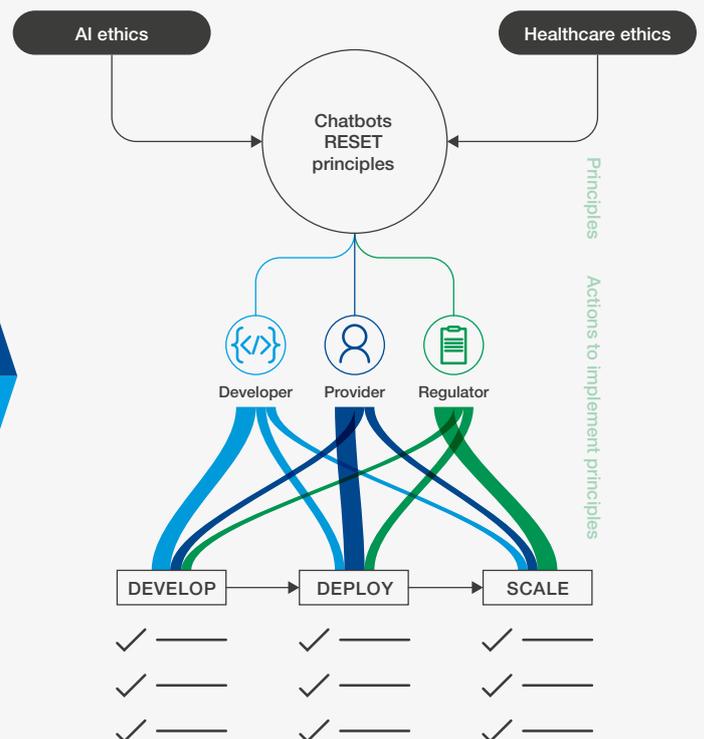
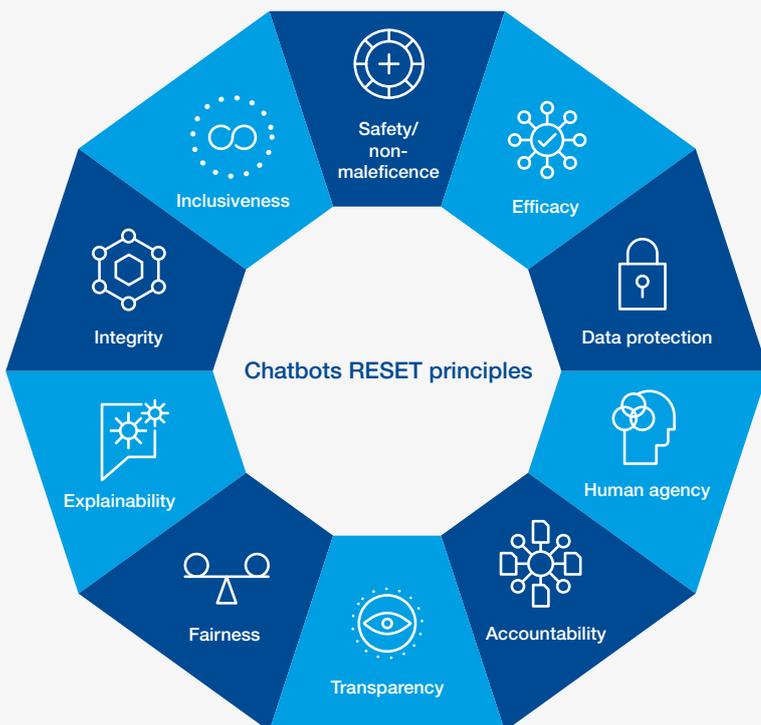
In the Chatbots RESET framework, each principle is further elaborated with a set of actions that developers, providers and regulators can take to implement the principle. The actions enable the various players using chatbots in healthcare to rapidly operationalize responsible behaviour.



The Chatbots RESET framework is lucidly articulated and very comprehensive; it has been built collaboratively with experts in the field.

Sujoy Kar, Chief Medical Information Officer, Apollo Hospitals, India

FIGURE 2 Principles and actions in the Chatbots RESET framework



2

Rwanda pilot: purpose and process

Using AI to perform patient triage is a powerful way to scale healthcare access when done responsibly.

Babylon is a global digital healthcare company providing services in Rwanda since 2016 through its partnership with the Government of Rwanda. Babylon's Rwandan entity, Babyl, allows Rwandans to consult with a doctor or nurse via any mobile device from anywhere in the country. Rather than travelling long distances to a health centre and waiting – sometimes hours – to be seen, patients can use Babyl to book appointments and receive a code via text for a prescription, referral or lab test. The service is integrated into Rwanda's National

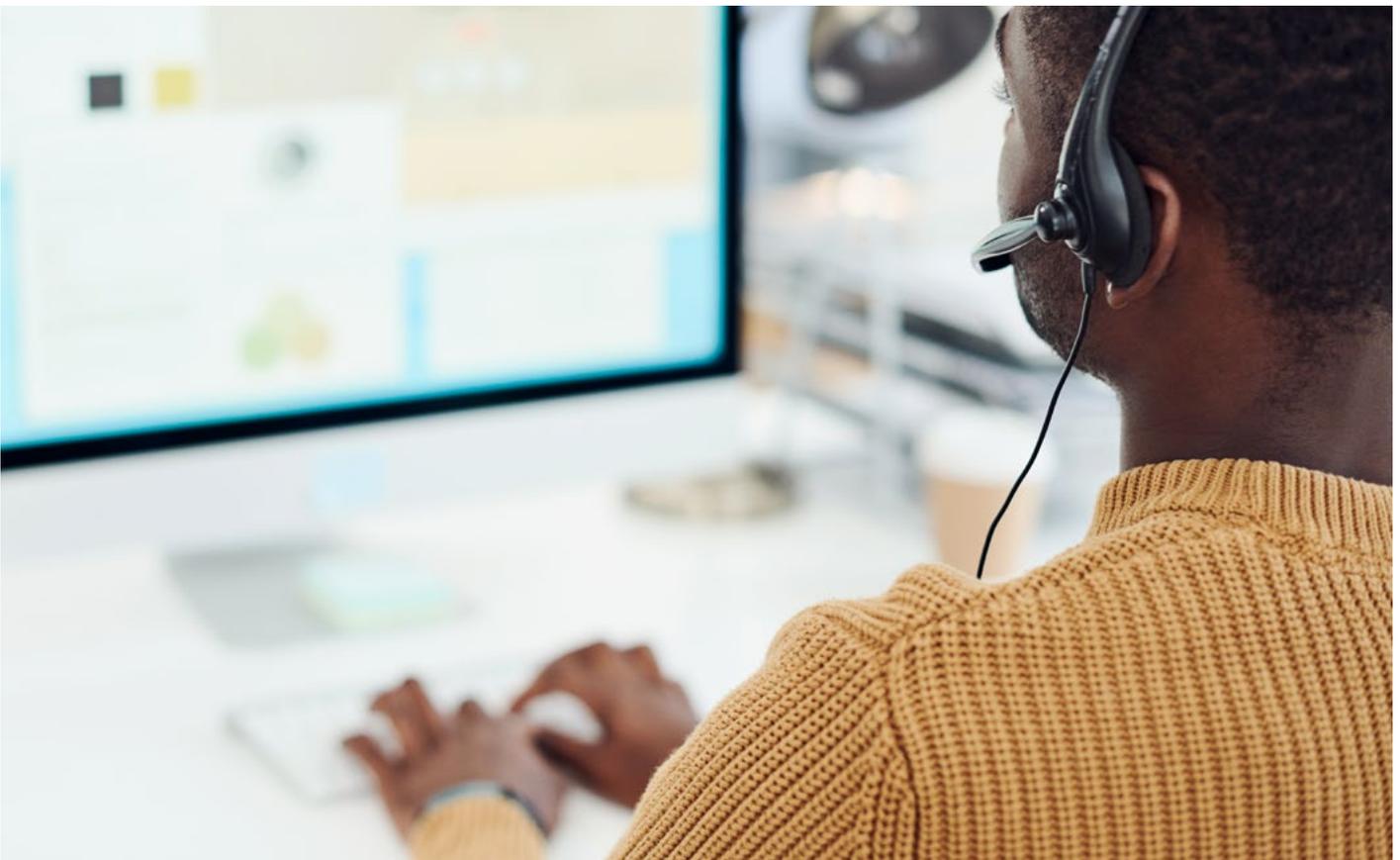
Health Insurance Scheme and is the largest digital health provider in the country, with over 2.6 million Rwandans signed up to use the service.

In 2021, Babyl introduced Babylon's AI-backed triage technology into this service for use by Babyl's call centre nurses when consulting with patients. The tool helps nurses ask patients the right questions, collect necessary information about a patient's symptoms, and provide them with insights to help choose the correct triage path.



Rwanda has been a pioneer in digital health and this pilot plays a vital role in promoting the responsible use of AI to further increase healthcare access.

Crystal Rugege, Managing Director, Centre for the Fourth Industrial Revolution Rwanda



2.1 Purpose of the pilot

A quick way to validate the Chatbots RESET framework is to assess the use of the framework principles and actions in terms of their relevance and practicality. Following the piloting of the framework in the context of four private sector organizations, it was necessary to understand the utility of the framework in a public healthcare scenario. Babyl's AI triage pilot offered an excellent opportunity to conduct this pilot.

Piloting goals

The goals of the pilot were to

- Demonstrate the achievement of project goals through the implementation of principles and actions from the Chatbots RESET framework
- Share the lessons from the pilot

These piloting goals serve to address broader goals: to maximize the beneficial uses of chatbots in healthcare and to minimize negative consequences.

2.2 Piloting process

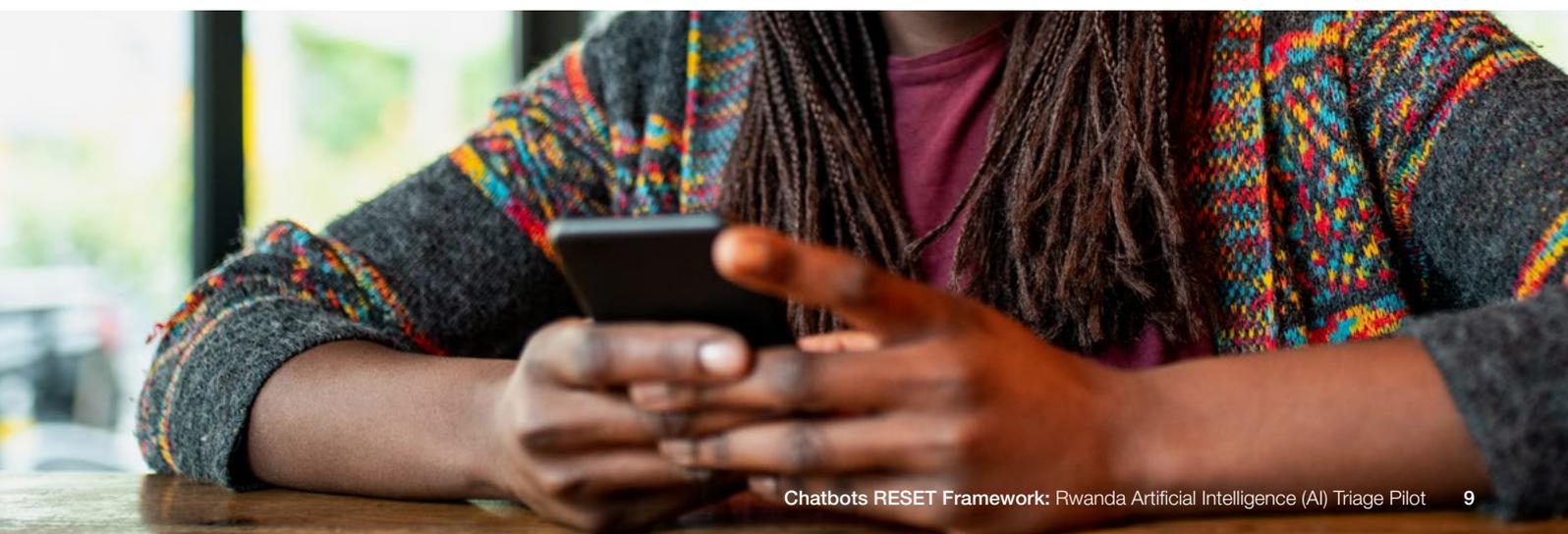
The Rwanda pilot was launched on 22 November 2021. During the following four months, the Forum's Artificial Intelligence and Machine Learning team members, Centre for the Fourth Industrial Revolution Rwanda members, and Babyl team members worked closely to select and implement three Chatbots RESET framework principles. During the pilot, a project community consisting

of medical, technology and regulation experts provided guidance on the implementation. The core team met fortnightly to review the status and plan the next steps. The project community met three times during the pilot to review and provide comments. Then a set of questions about the piloting experience were posed to Babyl.

BOX 1 Project community members

The project community that advised this pilot consisted of the following members:

- **Noella Bigirimana** (Deputy Director-General, Rwanda Biomedical Centre)
- **Maria Birkmyre** (Global Clinical Safety Officer, Babylon)
- **Kay Firth-Butterfield** (Head of Artificial Intelligence and Machine Learning, Centre for the Fourth Industrial Revolution, World Economic Forum)
- **Jean Baptiste Byiringiro** (Chief Digital Officer, Ministry of Health)
- **Patrick Gacirane** (Researcher/Analyst, AI Initiatives, Rwanda Utilities Regulatory Authority)
- **Ritah Kanyana** (Senior Strategy and Operations Lead, Babyl)
- **Sujoy Kar** (Chief Medical Information Officer, Apollo Hospitals)
- **Nupur Kohli** (Forum Global Shaper)
- **Nikhil Malhotra** (Chief Innovation Officer, TechMahindra)
- **Alain Ndayishimiye** (Centre for the Fourth Industrial Revolution Rwanda)



3

Rwanda pilot: lessons

Three Chatbots RESET principles were implemented in the Rwanda pilot.



This pilot has provided an opportunity to validate important ethical principles from the Rwandan context, informing our approach to responsible patient triage.

Shivon Byamukama, Managing Director, Babyl Rwanda

This section shares the experience and lessons from the pilot as a series of answers from Babyl and Babylon to questions about the pilot.

3.1 Why did you decide to pilot?

Like many countries worldwide, Rwanda struggles with healthcare accessibility and affordability issues while also facing ever-rising levels of chronic disease and severe healthcare workforce shortages. Integrating cutting-edge technologies, such as AI, into the existing Rwandan healthcare system is an important part of improving healthcare quality and accessibility throughout the country.

Babyl's service has been operating in Rwanda since 2016 and currently holds up to 4,000 consultations with patients every day. The integration of Babylon's AI-powered triage tool into this service was the next step in digitalizing the national healthcare system. Babyl's call centre nurses are using the AI-powered triage tool to help them work more efficiently and make better, quicker decisions for their patients. The tool has been fully localized for Rwanda and accounts for local language, epidemiology, culture and health system pathways.

The tool helps nurses ask patients the right questions, collect the necessary information about a patient's symptoms, and provide them with insights to help choose the correct triage path. If a follow-up appointment is needed, the patient information collected on the triage call is passed on to the doctor, saving time for both the nurse and the patient.

Babylon's technology has helped extend quality care to people all around the world, from hospitals in England's National Health Service (NHS) to some of the most rural corners of the United States and, through Babylon's partnership with major insurer Prudential, to more than 10 South-East Asian countries.

The Rwandan healthcare system presented a new opportunity for Babylon as it was the first time the company was launching AI in a country with low smartphone penetration. Traditionally Babylon's technology had always been accessible directly to users via a smartphone app. However, the situation in Rwanda meant that a new way was needed to make this technology accessible to everyone. Deploying this technology for use by Babyl's nurses in call centres meant that smartphone ownership was no longer a prerequisite to access this service, as users could call from any phone.

Babylon decided to do the pilot as it was an opportunity to evaluate its internal processes for the development of AI solutions against a framework designed by leading global experts that strongly aligns with Babylon's commitment to being a learning organization.

“ Integrating cutting-edge technologies, such as AI, into the existing Rwandan healthcare system is an important part of improving healthcare quality and accessibility throughout the country.

3.2 Which healthcare application did you pilot?

AI Triage Babylon Call Centre (AI triage tool) is a chat-based informational service operated by a medical professional (nurse) who enters the patient's reported medical concern as an initial input to the AI triage tool. The nurse then goes through a series of questions matched to that input. After this, the call centre nurse is provided with information about the location and timescale in which the patient might require treatment. The information provided by the service may or may not coincide with the professionally established information that the nurse ultimately decides to provide to the patient. The tool is intended to act as an informational service for the existing triage process, with the benefit of standardization of clinical practice and the product to be used solely by a qualified medical professional (nurse) who is the ultimate decision-maker on whether to accept the information generated by the service.

Types of conditions covered

The conditions for which the service can provide information to the medical professional include

those commonly encountered in primary care localized in Rwanda (excluding pregnancy and some mental health and dermatological conditions).

Intended user and patient population

The AI Triage Babylon Call Centre is intended to be operated by call centre-based qualified medical professionals (nurses). The patient population it ultimately serves includes cisgender people (people whose gender identity matches their sex at birth) presenting with symptoms of a common primary care condition (excluding pregnancy and some mental health and dermatology conditions).

The medical professional can use the product to generate information for patients of all age groups. Feedback is collected so that changes can be made to improve the system as required. Figure 3 presents the form the medical professionals complete so Babylon can review and analyse and, if needed, support changes to the product.

FIGURE 3 Triage outcome review screen

Triage Outcome Review

Is the triage outcome safe?

Yes

No

Is the triage outcome appropriate?

Yes

No

Do you have any more feedback? (optional)
Do not enter any personal identifiable information into this field (e.g. names, date of birth)

Close

Source: Babylon



3.3 Which AI technology did you use?

“ AI Triage Babyl Call Centre is a chat-based informational service operated by a nurse who enters the patient’s reported medical concern in the AI triage tool and then asks questions matched to that input.

The conversational AI functionality was provided to call centre nurses in Rwanda through the implementation of a localized version of Babylon’s AI triage tool. AI Triage Babyl Call Centre is a chat-based informational service operated by a medical professional (nurse) who enters the patient’s reported medical concern as an initial input to the AI triage tool and then goes through a series of questions matched to that input. After this, the call centre nurse receives some information about the location and timescale in which the patient might require further care.

Combining the successful history of Babylon and significant developments of the Babylon chat-based triage system, the current version in Rwanda is powered by a Bayesian network augmented by rules-based safety netting. It is designed to provide users (in this case, triage nurses) with triage advice alongside an explanation of why this action has been suggested; this consists of any reported symptoms that require urgent attention and a list of possible causes for the user’s symptoms.

It is based on a probabilistic graphical model (PGM) of primary care medicine that models the prior probabilities of diseases and the conditional dependencies between diseases, symptoms and risk factors.

Medical experts create the structure of the graph (meaning the connections between diseases, symptoms and risk factors) and it takes into account epidemiological data, including the incidence of diseases. The model is used to deduce the possible underlying diseases that cause the user-entered symptoms using Bayesian inference methods. This allows the AI-powered triage system to offer the most likely causes of the symptoms entered by a user and generate follow-up questions that provide information about the disease(s) under consideration. To guarantee the safe triage of patients with symptoms that require urgent treatment (regardless of their underlying causes), the model is augmented with a set of rules that dictate specific questions and triage actions when particular combinations of symptoms are present.

3.4 How was this AI technology developed for the pilot?

“ Like many countries worldwide, Rwanda struggles with healthcare accessibility and affordability issues while facing ever-rising levels of chronic disease and severe healthcare workforce shortages.

Triage is currently available in Canada, Cambodia, Hong Kong, Indonesia, Laos PDR, Malaysia, Myanmar, the Philippines, Singapore, Taiwan, Thailand, the United Kingdom, the United States and Viet Nam. Each model is specific for that particular country, using epidemiological data to define the model to ensure it is localized for that specific country. Babylon has developed its localization process in line with best-practice standards already available as the safety and performance of the product are paramount.

The AI triage tool originated from Check, released in mid-2016, and an updated release of the product known as Checkbase, released in mid-2017. This was followed by modifications to the AI triage tool and evolving releases known as Triage. This was to facilitate product enhancements continuously.

The latest version of the product was localized for Rwanda and delivered as part of the pilot. The intended use, user and population were updated to

reflect the localization work within Rwanda, including language to support the deployment of the AI Triage Babylon Call Centre. The pilot was to demonstrate that the Rwandan model is appropriate in its setting and with the view to a further roll-out to health centres. The ultimate aim is for it to be used by patients.

Babylon's Triage is done by triage nurses and is currently done for the Mutuelle patient base. Triage allows for the sifting of appointments to identify whether the presenting conditions are within the scope of service or not. After a Triage appointment, the nurse will either refer the patient to a physical facility if it is an emergency or out of scope or direct the patient to complete a full consultation with a general practitioner or senior nurse.

The medical professional can use the product to generate information for patients of all age groups. Additional standard operating processes were created and training was delivered to the nurses involved in the pilot.

3.5 What are the challenges and opportunities in your use of AI in this application and in Rwanda?

The AI triage tool now available to Babylon's nurses has been fully localized for Rwanda and accounts for local language, epidemiology, culture and health system pathways. Localization for any model is always a challenge. However, having robust processes in place that support governance ensures that product safety and performance can be measured appropriately. Babylon has had experience in delivering localized models to several countries and this process has been improved and has evolved and is in line with best practices.

Opportunities

1. Standardizing care, ensuring consistently high quality and accurate triage advice occurs with every encounter
2. Improving access and affordability – with the aim to make this directly available to the end-user as smartphone usage increases

3. Automating medical administration processes – all the information about symptoms and patient demographics recorded by the nurse can be passed on to the doctor

Challenges

1. Managing complex medical cases with multiple co-morbidities
2. Building trust about the use of AI for both healthcare professionals and users

Like many countries worldwide, Rwanda struggles with healthcare accessibility and affordability issues while facing ever-rising levels of chronic disease and severe healthcare workforce shortages. The introduction of Babylon's AI is a big step forward for the company's commitment to supporting the government's agenda to digitally transform the national healthcare system.



3.6 What was your approach to governance in this pilot?

Governance plays a vital role in ensuring users are not forgotten about at any time, including through software that users can download and that is informative, especially through systems using AI. Governance is crucial in deploying any software, whether through health and wellness apps or medical devices at the higher end of the spectrum. Having a governance system in place ensures alignment between safety and innovation such that they complement each other rather than act as a block on innovation. Any application that could potentially cause harm should have due diligence with complete

transparency and accountability embedded into the governance framework. This should be part of an organization's overall governance framework, of which the AI triage tool is one part.

Good end-to-end governance should cover the entire software development life cycle, from development to deployment within a live environment.

Below is a high-level description of the processes followed and supporting documentation created to support it.

Clinical safety plan

A clinical safety plan describes the risks associated with the proposed AI Triage Babyl Call Centre solution and how they will be derived and assessed. It describes the roles and responsibilities of the Risk Management Team members involved in this process as part of this release. The clinical safety plan also defines how risks associated with the functionality released will be managed throughout the application's lifetime and as such applies to the design, development, deployment and decommissioning of the applications included. It is reviewed and updated during the application's life cycle as necessary. This plan also considers engineering and security risks and operational risks to ensure end-to-end safety for all Babylon system users.

Additionally, two further clinical safety deliverables were produced as part of this activity: a Hazard Log and a Clinical Safety Case Report.

The clinical safety plan also specifies the approach for developing and maintaining these two documents. It clearly sets out the solution that is being delivered for Babyl Rwanda. Any core product/application changes will be managed as part of ongoing releases and relevant product teams. This will be the Triage Maintenance Release Process and the Rwanda Clinical Portal Maintenance Release Process.

Risk acceptance criteria have already been defined for the system. The following activities took place:

- New system hazard identification and update to existing hazards via a Hazard Analysis Workshop with risk management team members
- Following the secure software development life cycle
- Review of safety work to support Babyl Rwanda AI Triage Call Centre product release
- Review of new changes as part of the Babyl Rwanda AI Triage Call Centre product release
- Update of the Hazard Log to support Babyl Rwanda AI Triage Call Centre product release
- Review of Test Summary Report, which demonstrates end-to-end testing of the solution
- Operational readiness verification
- Safety case to support Babyl Rwanda AI Triage Call Centre product release

The above was carried out to support the call centre pilot with trained nurses.

“ Good end-to-end governance should cover the entire software development life cycle, from development to deployment within a live environment.

Test plan

A test plan was created and executed, and premarket testing was carried out involving engineering, quality assurance and clinical staff (doctors on the live register). This is part of the Babylon process to ensure that risk controls implemented are working as designed. Any failed tests are reviewed, and changes to the system are incorporated, testing is rerun, and case cards are updated as necessary. This is a continuous process. The Hazard Log identifying system hazards (meaning potential sources of harm) is updated accordingly.

All testing was performed in accordance with the AI Triage Babyl Call Centre Product Solution Release and Testing Strategy. Where additions were made, they have been discussed within the testing methodology.

Within direct scope for testing:

- AI Triage Babyl Call Centre – Back-end
- AI Triage Babyl Call Centre web app – Front-end

The clinical safety team has reviewed both test summary reports and assessed risks. Any known defects deemed unacceptable were fixed and retested before going live.

- Triage back-end pre-market testing
- Component and integration testing

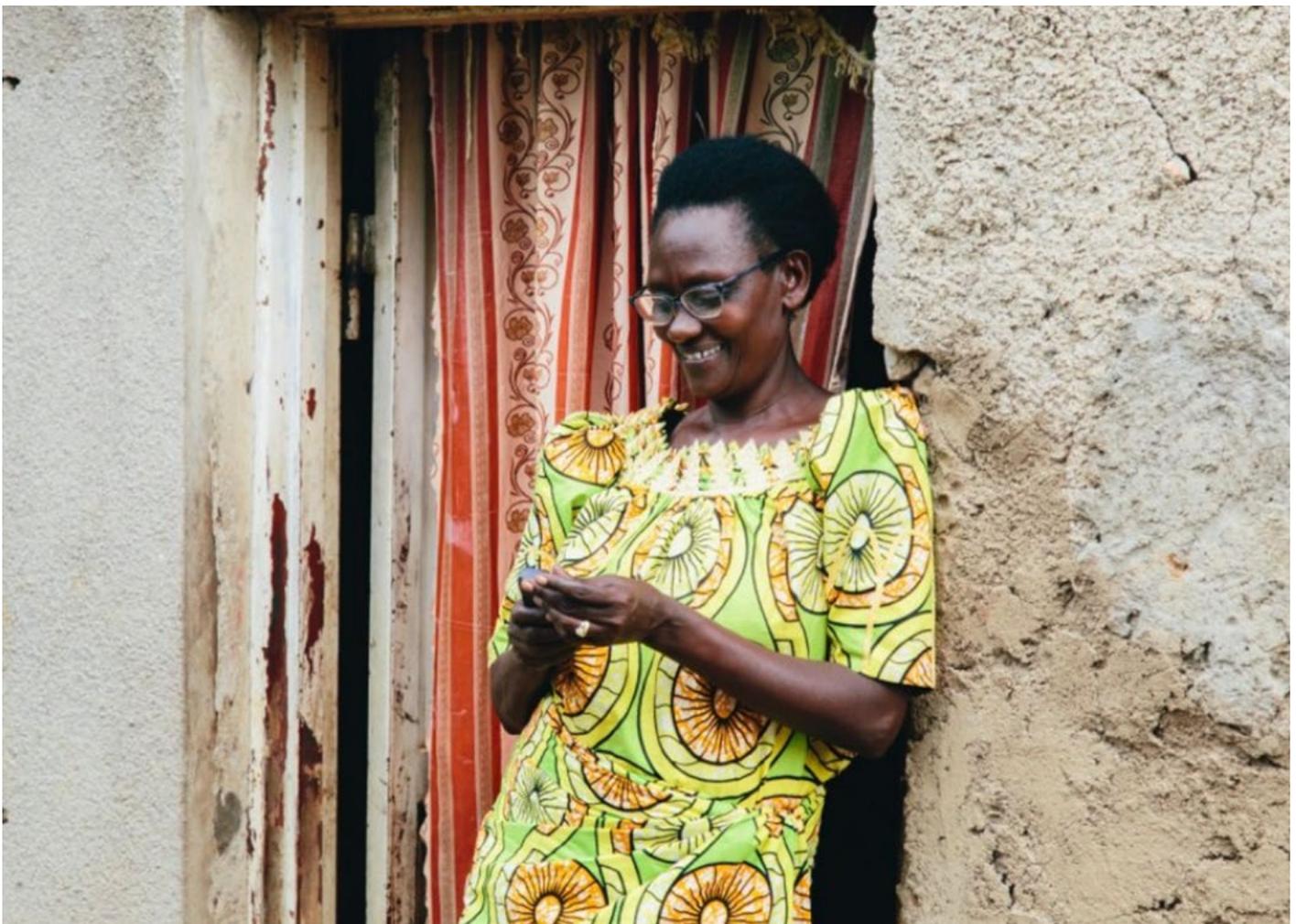
This has been completed on the back-end triage model, which concluded there were no failed tests. All tests have passed and are detailed in a Test Summary Report.

Clinical validation

Clinical validation of the localized triage model for Rwanda was also completed through auto-validation and clinical validation of the triage outcome copy. Regional doctors in Rwanda

concluded 7 rounds of clinical validation and 10 cases were flagged. All 10 cases were fixed before release; therefore, no flagged or outstanding defects were found after clinical validation.

Image credit: Babylon



“ Any application that could potentially cause harm should have due diligence with complete transparency and accountability embedded into the governance framework.

Product acceptance testing

This has been completed and 3 issues were identified and assessed for risks. These were deemed as being of a tolerable level and did not

impact patient safety. These were identified in the Clinical Safety Case for AI Triage Babyl Call Centre v1.0.

Triage front-end pre-market testing

Unit/integration testing

Unit and integration testing has passed with no issues identified to support going live.

User interface (end-to-end) testing

Completed with two failed tests that were at tolerable risk limits; these have been assessed for risks from a clinical safety perspective and did not impact patient safety.

User acceptance testing

User acceptance testing was completed with the clinical operations team in Rwanda. Four outstanding issues remain that are of tolerable risk limits. These have been assessed for risks from a clinical safety perspective and did not impact patient safety.

Additionally, as part of the deployment activities following release, monitoring occurred as part of the governance process. This assists with tracking and documenting mistakes attributable to the chatbot, which are shared with developers and regulators for continuous improvement.

Babylon/Babyl tracked these in two ways: (i) a feedback process via Freshservice (Babylon's incident management system), including complaints; (ii) feedback from users – complaints will be constantly monitored along with monitoring of the system in the live environment via post-release monitoring, a plan for which was created. Processes will also be created to support safety. They will include escalation and meetings to ensure the timely implementation of changes and the clinical safety assessment of each issue.

The following steps are followed when errors are encountered in the chatbot:

1. If the issue is due to the internet connection, the nurse informs the patient and then reverts to the manual triage flow
2. If the patient complains that the chatbot is taking too long, the nurse offers to proceed without AI
3. If the chatbot does not identify the symptom, the nurse reverts to the manual flow
4. If the chatbot freezes midway, the nurse terminates and reverts to the manual flow.

The following additional steps are taken:

- The nurse who encounters the above scenarios notifies the supervisor if the flow was abruptly interrupted before they could complete. The support clinician assesses the root cause, such as loss of internet connection, slow speed of flows, and system freeze.
- The nurse reports details of the feedback and product tickets are also raised for technical support.
- When the flow is interrupted, the nurse informs the patient and then reverts to the manual triage flow (without the AI tool).
- Every week, the team of supporting clinicians reviews all reported issues and sets up the necessary actions to have them resolved.

When feedback is received, such as flagged cases, these are assessed for safety and the Hazard Log is reviewed accordingly. If changes are needed, these are shared with the clinical, engineering and product teams and appropriate changes are made to the system and delivered in a maintenance release. Safety mitigations are in place as the user is a nurse and will have enough knowledge and experience to ensure patients are signposted and seen by the appropriate service based on need.

Monitoring, which is best practice, is used to support clinical risk management and ensure that the triage product in Rwanda is operating safely within its intended use. Additionally, it is a way to ensure that any new functionality released is not adversely affecting patient safety. The monitoring document outlines the daily monitoring activities following this first release of the triage product in Rwanda, a region in which Triage is not a medical device. Daily monitoring will launch alongside standard post-release monitoring as soon as the product goes live. The main objective of daily monitoring is to support the requirement that the product be appropriately safe post-release.



3.7 What is your governance process?

Patient safety is paramount at all times, along with clear accountability on roles and responsibilities and transparency of user safety. In this pilot, Babylon is the developer and relevant actions have been incorporated into the project development process to enable the responsible implementation of the Babylon chatbot; Babyl Rwanda is the healthcare provider and integrates actions from the framework in its workflows to promote responsible deployment.

Babylon has a clinical risk management process for clinical safety from a technology perspective that clearly defines the roles and responsibilities of the clinical safety team and product risk management team. This process has also been audited externally as part of the requirements of International Organization for Standardization (ISO) 13485 –

Medical devices – Quality management systems – Requirements for regulatory purposes. Additionally, if there is an adverse outcome from a clinical risk management perspective, this demonstrates how the global clinical safety officer (CSO) can go directly to Babylon's top management for support if required. Internal documentation also reflects the risk acceptability criteria applied within Babylon.

As part of Babyl processes, the medical delivery team manages clinical governance, led by the medical director. A governance meeting takes place monthly with the global governance team to track the implementation of the quality improvement plan and risk register.

Babyl operations are audited annually, after which a report is shared with leadership for action.

3.8 Which principles and actions from the Chatbots RESET framework did you choose? Why did you choose them?

Babylon is demonstrating the Chatbots RESET principles of safety, transparency and accountability as part of this pilot to support the implementation of the AI Triage Babyl Call Centre in Rwanda (see Figure 1 for definitions).

Babylon chose these specific principles because user safety is paramount and transparency and accountability come with safety. This was another means of applying a framework agreed by global experts, testing internal processes and applying any lessons to these as Babylon is committed to being a leading organization that promotes learning.

3.9 How did you implement the actions?

The tables below summarize actions taken by Babylon/Babyl to implement each of the principles. The “Framework action” column contains the

framework recommendation and the “Babylon/Babyl implementation” column describes the steps taken to implement each action.



Safety

Stage	Framework action	Babylon/Babyl implementation
Develop	Design a robust hands-off system for situations when AI fails	Clinical risk management has demonstrated that safety has been considered. This has been via an agreed safety approach based on the project’s scope and considering the potential clinical safety impact by the global clinical safety officer, using a Clinical Safety Management Plan, Clinical Safety Hazard Log and a Clinical Safety Case Report to support the pilot. It is evidence-based and includes contributions from internal subject matter experts.
	Install safeguards to identify abnormal behaviour and prevent manipulation	This was the production of a test plan executed to account for normal and abnormal behaviour. This was part of the Clinical Safety Hazard Log.
Deploy	Track and document mistakes attributable to chatbot; share with developers and regulators for continuous improvement	Babylon/Babyl tracked this in two ways: (i) a feedback process via Freshservice (Babylon’s incident management system), including complaints; (ii) feedback from users – complaints are constantly monitored along with monitoring of the system in the live environment via post-release monitoring, for which a plan was created. Processes were created to support safety, including escalation and meetings to ensure timely changes and the clinical safety assessment of each issue.
	While suggesting diagnostic/treatment options, consider issues related to patient safety	This was captured via the Clinical Safety Case and supporting Hazard Log produced before deployment. Babylon staff also followed global policies, including the Global Clinical Risk Management Processes and Global Technology Incident Management Process.
	Train all relevant personnel on when and how to intervene	Clinical personnel who participated in the pilot received training on: <ul style="list-style-type: none"> a. Running AI triage flows using the tool b. Handling exceptions, e.g. conditions out of scope, termination of flow, etc. c. Assessing the outcomes and channelling them appropriately d. Capturing and reporting feedback to the monitoring team (general practitioners).



Accountability

Stage	Framework action	Babylon/Babyl implementation
Develop	<p>Ensure that chatbot workflows are audited by a human at least every year internally to maintain current status and accuracy</p>	<p>Babylon used current processes already in place to measure safety and appropriateness. This happens regularly, usually every 3 months and via post-release monitoring if there is a major release. This happened as soon as it went live.</p>
	<p>Within the audit, be transparent about whether the chatbot is dynamically learning or is static</p>	<p>Babylon used current processes already in place to measure safety and appropriateness. This happens regularly, usually every 3 months and via post-release monitoring if there is a major release. This happened as soon as it went live. Any flagged cases were assessed for clinical safety and documented. Changes, if required, were delivered via a Maintenance Release, which had also been safety assured.</p>
	<p>Seek clinical inputs in the decision to implement a chatbot</p>	<p>Babylon has an established Global Clinical Risk Management Process that demonstrates that clinicians are involved and provide input. The clinical safety plan provided evidence of this as well.</p>
Deploy	<p>In case of “accidents”, provide a full explanation of why the chatbot did what it did</p>	<p>Babylon has well-established processes for staff to follow, as detailed in Babylon’s Global Technology Incident Management Process. All safety-related incidents are assessed, root cause analyses are recorded and, if required, changes are made.</p> <p>As the nurse’s role in using the chatbot is critical, it has been established that the nurse has to agree with the outcome of the triage, otherwise they have to use their own judgement and further orient the patient through the manual process. The nurse will then capture inappropriate outcomes as feedback within the tool for further investigation.</p>
	<p>Create a mechanism for accountability, especially if there is an adverse outcome</p>	<p>This is contained within Babylon’s clinical risk policy, which defines roles and responsibilities.</p> <p>Babyl has a clinical governance structure with clear roles and responsibilities. The medical director oversees all aspects within the division. Furthermore, the structure has a governance lead responsible for reporting, tracking and updating the risk register regularly. These are also escalated to senior leadership for further action and attention.</p>
	<p>Keep a comprehensive record of data governance</p>	<p>Babylon has an incident management system that is used to record all incidents. It includes clinical safety assessments for issues identified within the live environment.</p>



Transparency

Stage	Framework action	Babylon/Babyl implementation
Develop	<p>Use a chatbot persona clearly distinct from that of a human</p>	<p>Babylon uses user research to inform the design of the system. This feeds into the Hazard Log created for the AI triage tool and follows best practice standards, such as ISO Standard IEC 62366 on the application of usability engineering to medical devices.</p>
	<p>Explain decision or inference in plain language, infographic or video</p>	<p>Babylon demonstrated this via the Product Engineering Document (PED) for the AI triage tool. This formed part of the Clinical Safety Case report of how the AI triage tool works.</p>
	<p>Publish limitations of the chatbot (possibility of errors and consequences) and reliability of the chatbot</p>	<p>This was captured in Babylon documents such as warnings and labelling and will also be evident in the Clinical Safety Case produced.</p>
Deploy	<p>Do testing in realistic conditions (pilot) to verify that the chatbot can inform the user if there are issues with understanding</p>	<p>Internal testing covered this as call centre nurses, not patients, use the product. Additionally, premarket testing provided feedback via a test plan that will also have a supporting test summary report and was part of the Clinical Safety Case Report.</p>
	<p>Inform users if the matter is non-serious, serious or critical</p>	<p>In this deployment the users are qualified nurses, not patients. This is a mitigation technique in itself as the nurse has clinical knowledge and will ensure the patient is treated accordingly based on their presentation and will have access to appropriate clinical services. This was shown in the Clinical Safety Case.</p>
	<p>Share limitations of the use of algorithm and data</p>	<p>This was demonstrated via the Hazard Log and Clinical Safety Case. Babylon also undertook component failure mode analysis, which fed into the safety artefacts – Hazard Log and Clinical Safety Case.</p>
	<p>Be explicit in distinguishing between recommendation and information</p>	<p>Based on the outcome of the triage flow, a Babyl nurse can:</p> <ol style="list-style-type: none"> 1. Recommend the patient visit a nearby health facility if the outcome is an emergency or out of scope of treatment; 2. Assess the probable condition provided by AI, then inform the patient to complete a consultation. <p>This process was written up in the Standard Operating Process developed by the clinical leads and was part of clinician training for nurses.</p>

3.10 What is your feedback on the Chatbots RESET framework?

The Chatbots RESET Framework was helpful, informative, and easy to follow as the format is simple and the principles align with global best practices. The framework ensured consistency and structure, which facilitated the identification of any gaps. The experience of using the framework in a real-world setting showed that developers and those deploying AI software can easily adopt it and support regulators when considering the regulatory framework for AI. The Chatbots RESET framework was beneficial in evaluating existing governance processes for the implementation of a new initiative – the AI Triage Babyl Call Centre (AI triage tool). An example was a renewed look at accountability and ensuring the correct people with the right knowledge and skillsets were available.

Additionally, Babylon used the framework as a tool to review their current processes and align these with published best practices in the public domain, such as ISO 14971:2019 on the application of risk management to medical devices. This helped ensure risk management is applied throughout the development and deployment stages of releasing the AI software in Rwanda. Other principles from the Chatbots RESET were reviewed and will be assessed and built into Babylon's current governance processes. The framework addresses critical areas to be considered in using chatbots and AI. Applying the framework by adopting fundamental principles can help further ensure that end-to-end governance is achieved.

3.11 Would you like to share any best practices?

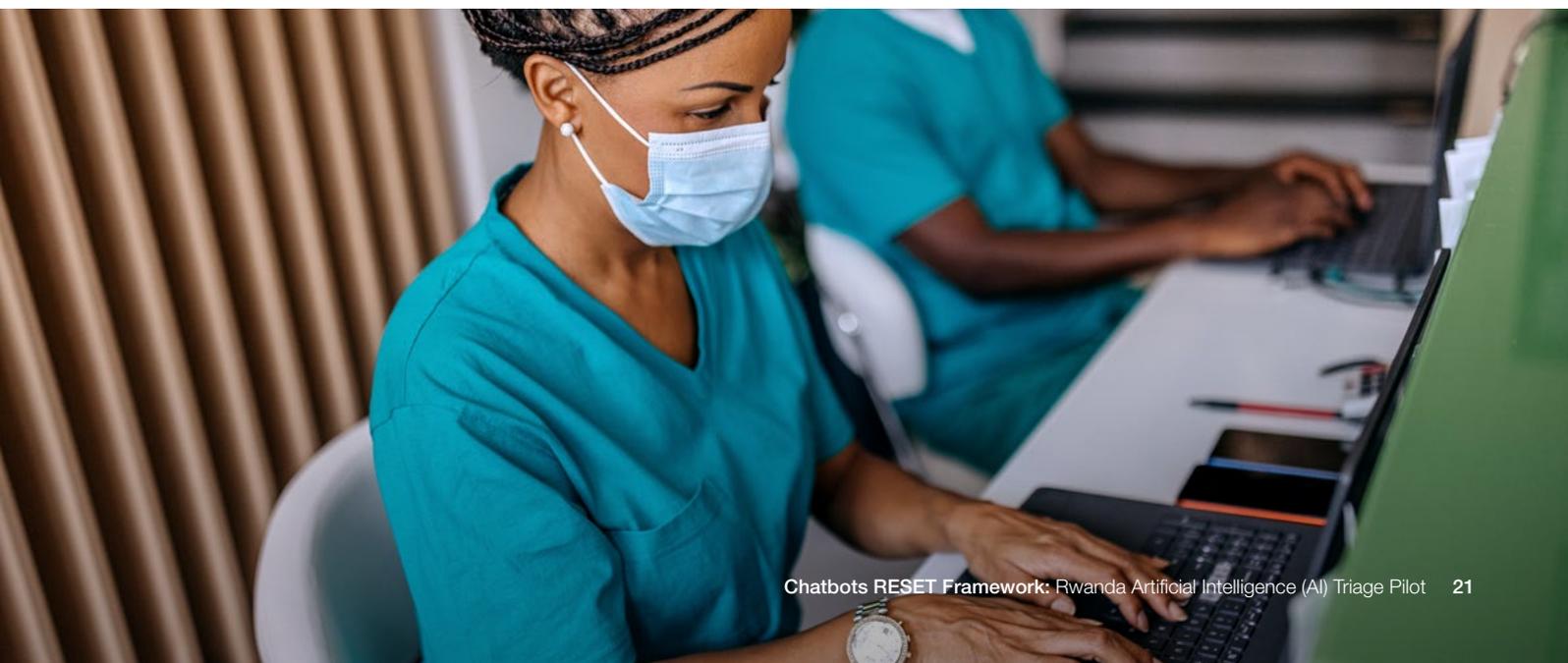
The key to this pilot is ensuring that current best practice widely available in the public domain is actually used and built into existing processes. This should include the developer and the organization that is deploying any software being used in a healthcare setting that can potentially cause harm. Following the best practice and applying it to the pilot ensured that the pilot was a success. Additionally, a feedback loop now ensures that improvements can be adapted and continual risk management is applied by standards such as:

- BS EN ISO 14971 Medical Devices: Application of Risk Management to Medical Devices
- DCB 0129 Health Informatics – Application of clinical risk management to the manufacture of health software
- DCB 0160 Health Informatics – Application of clinical risk management to the deployment of health software



As pioneers in developing and deploying AI technology in healthcare, it has been highly rewarding for Babylon's governance approach to be validated by leading global experts who developed this superb framework.

Mobasher Butt, Global Medical Vice-President, Babylon



Conclusion

Piloting outcomes could lead to the broader adoption of AI in healthcare.

This report shares the piloting experience of implementing the Chatbots RESET framework in a healthcare call centre scenario in Rwanda. The knowledge and lessons learned from this pilot and the preceding four pilots reinforce the benefits of using frameworks like Chatbots RESET as “soft governance” mechanisms for regulating AI instead

of regulations, which trail behind. The hope is that the positive outcomes of these pilots will inspire other private and public sector organizations to adopt AI to increase healthcare access for populations by using the principles and actions from the Chatbots RESET framework to achieve the responsible use of the technology.

Contributors

World Economic Forum

Arunima Sarkar

Project Lead, Artificial Intelligence
and Machine Learning

Venkataraman Sundareswaran

Mitsubishi Chemical Holdings Fellow,
Artificial Intelligence and Machine Learning

Babylon

Maria Birkmyre

Global Clinical Safety Officer

Mobasher Butt

Global Medical Vice-President

Centre for the Fourth Industrial Revolution Rwanda

Joris Cyizere

Strategy Lead

Crystal Rugege

Managing Director

Endnotes

1. *Chatbots RESET Framework Pilot Projects: Using Chatbots in Healthcare*, World Economic Forum, 2021.
2. *Chatbots RESET: A Framework for Governing Responsible Use of Conversational AI in Healthcare*, World Economic Forum, 2020.



COMMITTED TO
IMPROVING THE STATE
OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org