

1 **Natural language processing to evaluate texting conversations between patients**
2 **and healthcare providers during COVID-19 Home-Based Care in Rwanda at scale**

3

4 Richard T Lester^{1¶*}, Matthew Manson^{1¶}, Muhammed Semakula^{2,3}, Hyeju Jang^{4,5}

5 Hassan Mugabo³, Ali Magzari⁶, Junhong Ma Blackmer⁷, Fanan Fattah¹, Simon Pierre

6 Niyonsenga³, Edson Rwagasore³, Charles Ruranga⁸, Eric Remera³, Jean Claude S.

7 Ngabonziza^{3,9}, Giuseppe Carenini⁵, Sabin Nsanzimana²

8

9 ¹Division of Infectious Diseases, Department of Medicine, University of British Columbia,
10 Vancouver, British Columbia, Canada

11 ²Rwanda Ministry of Health, Kigali, Rwanda

12 ³Rwanda Biomedical Centre, Kigali, Rwanda

13 ⁴Luddy School of Informatics, Computing, and Engineering, Department of Computer
14 Science Indiana University Indianapolis, Indianapolis, Indiana, United States

15 ⁵Department of Computer Science, Faculty of Science, University of British Columbia,
16 Vancouver, British Columbia, Canada

17 ⁶Department of Electrical and Computer Engineering, University of British Columbia,
18 Vancouver, British Columbia, Canada

19 ⁷Department of Mathematics, University of British Columbia, Vancouver, British
20 Columbia, Canada

21 ⁸African Center of Excellence in Data Science, University of Rwanda, Kigali, Rwanda

22 ⁹Department of Clinical Biology, University of Rwanda, Kigali, Rwanda

23

24 *Corresponding author

25 Email: [rl Ester@mail.ubc.ca](mailto:rl Lester@mail.ubc.ca)

26

27 ¶ These authors contributed equally to this work.

28 **Abstract**

29

30 Isolation of patients with communicable infectious diseases limits spread of pathogens
31 but can be difficult to manage outside hospitals. Rwanda deployed a digital health
32 service nationally to assist public health clinicians to remotely monitor and support
33 SARS-CoV-2 cases via their mobile phones using daily interactive short message
34 service (SMS) check-ins. We aimed to assess the texting patterns and communicated
35 topics to understand patient experiences. We extracted data on all COVID-19 cases
36 and exposed contacts who were enrolled in the WeTel text messaging program
37 between March 18, 2020, and March 31, 2022, and linked demographic and clinical
38 data from the national COVID-19 registry. A sample of the text conversation corpus was
39 English-translated and labeled with topics of interest defined by medical experts.
40 Multiple natural language processing (NLP) topic classification models were trained and
41 compared using F1 scores. Best performing models were applied to classify unlabeled
42 conversations. Total 33,081 isolated patients (mean age 33.9, range 0-100), 44%
43 female, including 30,398 cases and 2,683 contacts) were registered in WeTel.
44 Registered patients generated 12,119 interactive text conversations in Kinyarwanda
45 (n=8,183, 67%), English (n=3,069, 25%) and other languages. Sufficiently trained large
46 language models (LLMs) were unavailable for Kinyarwanda. Traditional machine
47 learning (ML) models outperformed fine-tuned transformer architecture language
48 models on the native untranslated language corpus, however, the reverse was
49 observed of models trained on English-only data. The most frequently identified topics
50 discussed included symptoms (69%), diagnostics (38%), social issues (19%),

51 prevention (18%), healthcare logistics (16%), and treatment (8-5%). Education, advice,
52 and triage on these topics were provided to patients. Interactive text messaging can be
53 used to remotely support isolated patients in pandemics at scale. NLP can help
54 evaluate the medical and social factors that affect isolated patients which could
55 ultimately inform precision public health responses to future pandemics.

56

57 **Author Summary**

58

59 We present the first application of NLP for categorizing text messages between patients
60 and healthcare providers within a nationally scaled digital healthcare program. This
61 study provides unique insights into the circumstances of home-based COVID-19
62 patients during the pandemic. Our trained topic classification models accurately
63 categorized topics in both English and African language texts. Patients reported and
64 discussed both medical and social issues with public healthcare providers. This
65 approach has the potential to guide precision public health decisions and responses in
66 future outbreaks, pandemics, and remote healthcare scenarios.

67 **Introduction**

68

69 The practice of isolating patients with communicable infectious diseases reduces
70 transmission by separating infected individuals from healthy ones, and isolation at home
71 allows healthcare systems to focus on managing severe cases in hospitals.¹ However,
72 isolated patients may have changes to their condition that require attention from medical

73 professionals and can experience additional social stressors.² Isolation protocols have
74 even been considered a risk factor for mortality.³ Understanding the complexity of
75 facility and home-based isolation remains a critical consideration in effective pandemic
76 response planning.

77
78 At the pandemic outset, the Rwanda government COVID-19 Joint Task Force adopted a
79 short message service (SMS)-based digital health service, WeTel, to remotely monitor
80 and support community isolated cases and contacts of SARS-CoV-2 from the
81 centralized national command post. Its use continued throughout the emergency phase
82 of the pandemic as the key to their Home-Based Care (HBC) program.⁴ The WeTel
83 platform was selected based on its history of use in HIV programs, prior clinic evidence,
84 and importantly, its accessibility.⁵⁻⁷ Internet based tools and smartphone apps would
85 have been insufficient since, as a lower-middle income country (LMIC), only 23% of the
86 Rwanda population has internet-access; yet at least 90% of the population has access
87 to cellular phones.^{8,9} As in other applications of the WeTel platform, throughout the
88 COVID-19 response in Rwanda, automated 'check-in' messages were sent daily to
89 isolated patients, to which they could respond to report on their status and ask
90 questions. A manually performed topic classification of WeTel text messaging
91 conversations with patients during HIV care previously demonstrated that these open-
92 ended text messaging conversations contain wide ranging issues experienced by
93 patients that could be acted upon.¹⁰

94

95 Advances in natural language processing (NLP) enables rapid computational topic
96 analysis (classification and modelling) which can delineate large amounts of text into
97 topics of interest.¹¹ Topic classification models using machine-learning (ML) have been
98 used in clinical settings to label text messages exchanged between healthcare
99 providers and outpatients, but never to our knowledge in a national scale program.^{12–15}
100 Unfortunately, the uneven distribution of global healthcare data has the potential to
101 leave entire continents behind in the development of artificial intelligence (AI) for
102 healthcare.¹⁶

103
104 In this study, we aimed to analyze the complete corpus of SMS conversations between
105 Rwanda’s public health clinicians and isolated COVID-19 patients and contacts during
106 the first two years of the COVID-19 pandemic. We considered a clinical, social, and
107 logistical perspective using topic classification to better delineate the needs and
108 experiences of patients and the advice they were given, during isolation – a large-scale
109 approach that could be used to guide decision-making in outbreak management.

110

111

112 **Results**

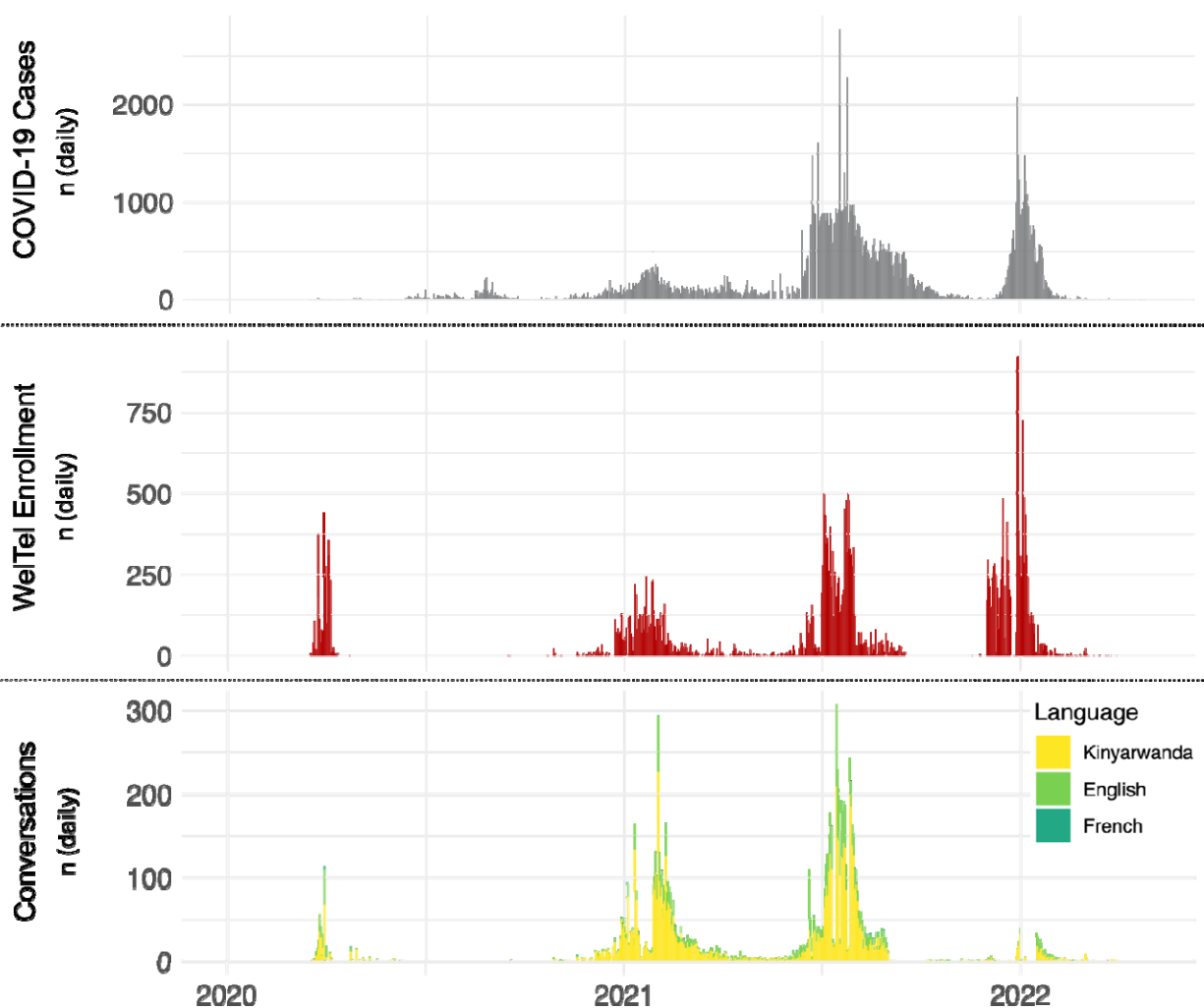
113

114 **Participation in text messaging conversations**

115

116 During the study period, Rwanda reported 131,190 cases of COVID-19, of whom
117 33,081 individuals (25%, 30,398 cases, 2,683 contacts) were registered in the WelTel

118 texting service. The numbers of patients registered, and the occurrence of text-message
119 conversations (with at least 3 interactive messages), appeared in waves similar to the
120 COVID-19 incidence reported in the national registry (Fig. 1).



121
122 **Figure 1. COVID-19 case incidence, and enrollment and use of the WelTel SMS text**
123 **messaging service during the COVID-19 pandemic in Rwanda.** *The grey plot captures the*
124 *four major waves of case incidence in Rwanda. The red plot captures enrollment of COVID-19*
125 *patients (cases and contacts) in the WelTel system throughout the study period. Between June*
126 *and September 2020, patients were not being enrolled in WelTel as the Rwandan government*
127 *transitioned to government servers to host the software. The yellow and green plot captures the*

128 number of patient-provider SMS conversations throughout the study period; most conversations
 129 occurred in Kinyarwanda (67%) and English (25%) and other regional languages.

130
 131 Of those registered, 6,021 patients (18%) used the texting service to generate 12,119
 132 conversations (Table 1). Controlled for other demographic and clinical factors, the odds
 133 of generating a conversation among females compared to males was 0.83 (95%CI:
 134 0.78-0.89; $p < 0.001$). Age did not significantly influence the frequency of conversations
 135 (OR per year increase in age = 1.00, 95%CI: 0.99-1.00, $p = 0.17$). Contacts were more
 136 likely to generate conversations than cases (OR 1.83 95%CI: 1.63-2.06, $p < 0.001$). The
 137 proportion registered patients texting conversations varied across waves of the
 138 pandemic (Wave one: 12%; two: 37%; three: 25%; and four: 3%).

139
 140 **Table 1. Characteristics of COVID-19 patients (cases and contacts) and usage of the**
 141 **WelTel text messaging service in Rwanda between March 18, 2020, and March 31, 2022.**
 142 *Percentages are within group. A proxy is an individual who allowed another/other individual(s)*
 143 *(e.g., a non-phone owner) to register for WelTel using their phone number (e.g., a mother could*
 144 *be a proxy for her non-phone-owning daughter). Shared refers to the individuals using a proxy*
 145 *(e.g., the daughter).*

	People			Conversations		
	COVID-19 CASES	ENROLLED	PRODUCING CONVERSATIONS	CONVERSATION LENGTH		
				Median # Messages	IQR	Range
TOTAL	131,190	33,081	6,021	3	2-5	2-44
COVID-19 Status (at registration)						
Case	131,190 (100%)	30,398 (92%)	5,322 (88%)	3	2-5	2-44
Contact	-	2,683 (8%)	699 (12%)	4	3-8	2-30

Location of Residence						
Kigali City	42,152 (32%)	33,018 (97%)	5855 (97%)	3	2-5	2-44
Northern	20,347 (16%)	111 (0.3%)	22 (0.4%)	4	3-6	2-16
Southern	25,814 (20%)	370 (1%)	60 (1%)	3	2-5	2-18
Eastern	18,007 (14%)	233 (1%)	25 (0.4%)	5	3-7-5	2-11
Western	24,870 (19%)	161 (0.5%)	15 (0.2%)	3	2-4-5	2-16
Unknown	-	188 (0.6%)	44 (0.7%)	4	2-9	2-19
Sex						
Male	63,637 (49%)	17,102 (52%)	3,385 (56%)	3	2-5	2-36
Female	67,545 (51%)	14,665 (44%)	2,608 (43%)	3	3-5	2-44
Unknown	8 (0.0006%)	1,314 (4%)	28 (0.5%)	4	3-7	2-14
Age (mean (SD, range))	36.2 (17.36, 0-115)	33.9 (14.51, 0-100)	34.0 (13.27, 0-91)	-	-	-
Proxy/shared	-	1,726/2,749	453 (-)	-	-	-

146

147 The median number of conversations per patient was one (IQR: 1-2, range: 1-18), and
 148 the median conversation length (# of discrete text messages) was three (IQR: 2-5,
 149 range: 2-44). Most conversations occurred in Kinyarwanda (8,183, 67%) or English
 150 (3,069, 25%), however other regional languages (French, Kiswahili, Ganda,
 151 Chewa/Nyanja) were also identified (<10%). Language proportions were similar to the
 152 proportion of the population considered literate in those languages in Rwanda.²⁷

153

154 **Comparison of NLP model performance**

155

156 The NLP models tested classified most topics with a level of performance above our
 157 cut-off F1-score (Table 2). The traditional ML models performed better than fine-tuned
 158 transformer models on untranslated native language conversation data. Best performing
 159 models for each topic met the pre-determined F1-score cutoff of 0.7 for six of the nine

160 topics meeting the sample size threshold of 100. Model performance was roughly
 161 correlated with the number of occurrences (within sample) of the given topic.

162
 163 **Table 2. Performance of classification models on original language (untranslated),**
 164 **and English-translated conversations for each topic of interest.** *Blue and orange fill*
 165 *indicate traditional-ML, and transformer architecture language models, respectively. Bold text*
 166 *indicates an F1 score meeting our performance cutoff (≥ 0.7). Topics are ordered by best-*
 167 *performing model F1 performance score (descending) when trained with original language*
 168 *(untranslated) conversations as ultimately used to label the full corpus.*

Topic (n)	Untranslated Conversations		English-translated Conversations	
	Classifier (Feature Extraction Method)	Performance (Precision Recall F1)	Classifier (Feature Extraction Method)	Performance (Precision Recall F1)
Diagnostic Methods (1030)	Random Forest (Character n-gram + TF-IDF)	0.93	BERT (NA)	0.95
		0.90		0.97
Symptoms (2002)	KinyaBERT (NA)	0.87	BERT (NA)	0.92
		0.94		0.96
Treatment (174)	KinyaBERT (NA)	0.85	Longformer (NA)	0.96
		0.88		0.96
Prevention (459)	Ridge (Character n-gram + TF-IDF)	0.77	Longformer (NA)	0.97
		0.84		0.93
		0.92		0.96
		0.91		0.94
		0.90		0.96
		0.81		0.95

Social (406)	Logistic Regression (Character n-gram + TF-IDF)	0.67	Longformer (NA)	0.88
		0.74		0.82
		0.70		0.85
Healthcare Logistics (281)	Logistic Regression (TF-IDF)	0.62	Ridge (Character n-gram + TF-IDF)	0.64
		0.81		0.88
Service Quality (425)	Ridge (Character n-gram)	0.69	Longformer (NA)	0.83
		0.67		0.77
		0.68		0.80
Lifestyle/ Behavioural (156)	Logistic Regression (TF-IDF)	0.71	Longformer (NA)	0.92
		0.65		1.00
		0.68		0.96
Technical/ IT (185)	Logistic Regression (Non-Binary Bag of Words)	0.80	BERT (NA)	0.83
		0.43		0.54
		0.56		0.65

169

170 In our language experiments, fine-tuned transformer models generally performed better
 171 when using the English (translated) training and testing corpus, exceeding the F1
 172 threshold in all but one topic.

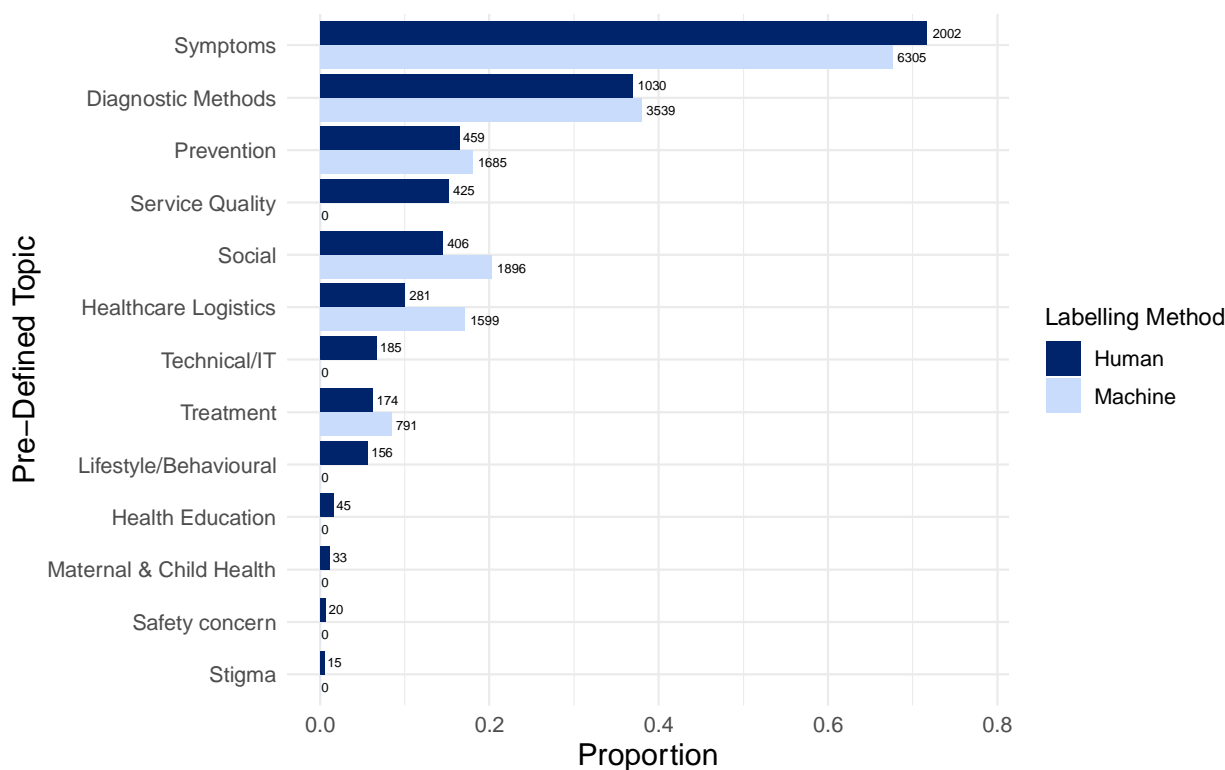
173

174 **Topic analysis of conversations**

175

176 Human labeling (training corpus) and machine topic labeling (full corpus) identified the
 177 topics and subtopics of interest at roughly similar frequencies (Figure 2 and S3 Fig).

178 Overall, medical topics were discussed more frequently than non-medical topics. The
 179 most frequent topic was symptoms (69%), the majority reporting a lack of symptoms
 180 (59%). Diagnostic methods (38%, e.g., receiving COVID-19 test results), prevention
 181 (18%, including non-pharmaceutical such as how to maintain isolation, and
 182 pharmaceutical such as vaccines), healthcare logistics (16%, e.g., asking where to get
 183 tested), and treatment (8%, including management of symptoms). Social topics (19%)
 184 included mention of culture or religion (5.5%), followed by discussing friends and family
 185 (4.9%). Manually labeled training data revealed important issues such as nutrition and
 186 food security (3.7%). Examples of text conversations are provided in Table 3.
 187



188
 189 **Figure 2. Proportion of conversations containing each topic of interest within the**
 190 **complete corpus (n=12,119) based on both human- (n=2,791) and machine-**

191 **labelling (n=9,328).** *Note: conversations were often labelled with multiple topics/subtopics*
 192 *therefore proportions do not sum to 1. Some topics did not have sufficient data to produce*
 193 *models, or the models developed did not meet our performance cutoff of $F1 \geq 0.7$. For these, the*
 194 *proportion estimates are based upon human labelling alone.*

195
 196 Comparing the odds of topic discussion within the full corpus by patient demographic
 197 and clinical factors revealed minor differences that are reported in (S4 Fig).

198
 199 **Table 3. Illustrative text conversations excerpts between public health officials**
 200 **and patients (cases and contacts) during the COVID-19 pandemic in Rwanda. “...”**
 201 *indicates excluded text for brevity. Message types include: ‘(S):’ – a ‘system’ message*
 202 *containing customizable content with automated sending; ‘(P):’ – a ‘patient’ message containing*
 203 *patient remarks in free text form; and ‘(C):’ – a ‘clinician’ message containing clinician remarks*
 204 *in free text form.*

CONVERSATION TOPIC	EXAMPLE
Reporting a lack of symptoms	<p><i>(S): Hello, how are you today? Do you have any symptoms of coronavirus? [alongside Kinyarwanda and French translations of the message]</i></p> <p>(P): We are ok. We do not have symptoms.</p> <p><i>(C): Hello! We inform you that you were tested negative for coronavirus. Thank you.</i></p>
Reporting symptoms	<p><i>(S): Are you OK (Yes/No)? Or tell us if you have new symptom [alongside Kinyarwanda translation of the message]</i></p> <p>(P): Good morning. I am coughing and I have flu but my situation is not alarming. I do not think that there is a problem.</p>

	<p>(C): Thank you for informing us. Explain your situation to the Doctor who is in charge of calling you. He is going to call you soon.</p>
Diagnosis	<p><i>(S): haven't heard from you yet. how are you? [alongside Kinyarwanda translation of the message]</i></p> <p>(P): Good morning. My name is #NAME#. I have been suffering from COVID-19 for days, but yesterday I went at #HEALTH FACILITY# to have a medical test for checking. Till now I have not yet got results via an sms. Can you help me to get those results if possible?</p> <p>(C): Good morning. We will inform people in charge to help you. Be patient.</p> <p>(C): #NAME# please note that you tested Negative on COVID-19 Ag RT on 2021-01-25. Code :*****</p>
Treatment	<p><i>(S): Check-in message</i></p> <p>(P): "...I took a Doliprane. Because I have little fever. 37.4..."</p> <p>(C): "Nice to hear that you are doing well. Doliprane helps for fever, body aches and headache..."</p> <p>(P): Yes, Thank you! Have a nice day</p>
Pharmaceutical prevention	<p><i>(S): Check-in message</i></p> <p>(P): "...Can you tell me where people suffering from heart issues and other diseases can go to get a vaccine?..."</p> <p>(C): "...Where do you usually go for a medical follow up...?"</p> <p>(P): "...#HEALTH FACILITY#"</p> <p>(P): I suffered from Covid 19 a month ago but now I am healthy.</p> <p>(C): You can contact them in order to know what they have scheduled for you.</p> <p>(P): Thank you</p>
Non-pharmaceutical	<p>(P): "...[Our] child has already tested positive...I have some</p>

<p>prevention</p>	<p>questions...Will not I be infected again once I share the same room and equipments with my child?...Is it possible to get a medical test for checking before ending 14 days of quarantine?..."</p> <p>(P): "...I am still suffering from coughing while the medicine I was taking is over..."</p> <p>(C): "...It is a must to go for a medical test for checking after 14 days. It would be better to clean your hands neatly, and to put on mask accordingly before taking your child..."</p> <p>(P): Thank you very much.</p>
<p>Social – food and nutrition</p> <p>Social – culture and religion</p>	<p>(P): "...I am suffering from COVID-19 and I live alone. I am starving and I don't have food. What can you help me? Thank you for your helpful answer."</p> <p>(C): Good afternoon. You can call the head of your village for a help. Food aid is being distributed these days. Thank you.</p> <p><i>(S): Check-in message</i></p> <p>(P): We are well this morning. We are really under the protection of Jesus.</p> <p>(C): That is good if you are well. Have a nice day.</p>
<p>Maternal & child health</p>	<p><i>(S): Check-in message</i></p> <p>(P): Hello. I am pregnant and I have tested positive for COVID-19. At the beginning I did not have any symptoms, but now I have many and I need your advice.</p> <p>(C): "...Be patient, you are going to get our support via a telephone call."</p>
<p>Gratitude</p>	<p>(C): Hello, do you have any health or COVID related concerns you would like to share with us today?</p> <p>(P): I tested negative for Covid 19. Thank you.</p>

	<p>(P): Thank you for your regular care.</p> <p>(P): Your advices are very important and we will always resort to them.</p>
Safety concern	<p><i>(S): Check-in message</i></p> <p>(P): We are in a very bad situation. ...We are mixed with people who have symptoms and those who do not have them, and we share the same bathroom and shower. Please, help us to get out of this situation before being infected.</p> <p><i>(C): Thank you for informing us. We are following up your situation. Keep on informing us. Thank you.</i></p>
Proxy example	<p><i>(S): Check-in message</i></p> <p>(P): Good morning. All of us, including children, are well. We don't have a problem. We hope to recover soon. Have a nice day. Thank you.</p> <p><i>(C): Thank you and have a nice day too.</i></p>

205

206

207 **Methods**

208

209 **Setting and digital health monitoring technology**

210

211 Rwanda, a Central African nation with a population of 14 million, was internationally
212 recognized for its prompt response to the COVID-19 pandemic.¹⁷ Their public health
213 command implemented the WelTel text messaging service [www.weltelhealth.com] on
214 March 18, 2020, coincident with detecting its first COVID-19 cases. This tool enabled

215 public health staff to remotely monitor and support individuals who tested positive for, or
216 were at high risk of contracting, SARS-CoV-2, and remained operational until the public
217 health emergency was officially declared over.

218
219 The WeTel platform is a secure web-based application that healthcare providers can
220 access from any internet-connected device. It interfaces with cellular networks enabling
221 it to send both pre-set and manually input SMS messages to registered mobile phone
222 users. It features a conversation dashboard that visualizes responses.

223
224 Following the initial facility-based quarantine of COVID-19 cases identified at border
225 entry points, Rwanda launched its HBC program when community transmission became
226 widespread. Under this program, mild cases and contacts were asked to stay home and
227 be remotely monitored. Patients with significantly worsening clinical status were triaged
228 to hospitals. Individuals were enrolled as 'cases' if they tested positive for SARS-CoV-2
229 or as 'contacts' if they were deemed high risk due to exposure. Patients without
230 personal cell phone numbers were registered via a household member, friend, or
231 neighbor.

232
233 The messaging protocol for COVID-19 monitoring involved sending bulk automated
234 daily 'check-in' texts in Rwanda's principal languages (Kinyarwanda, English, French)
235 daily to registered patients throughout their isolation period. Patients could respond via
236 text in their own words to indicate their status and/or ask questions. Replies indicating a
237 problem or question were flagged for follow-up. Public health clinicians responded to

238 flagged messages to provide information or advice, forming interactive text

239 conversations.

240

241 **Included datasets and description**

242

243 Patient registration and all text communication data generated between March 18, 2020

244 and March 31, 2022 was exported from the WelTel database and linked to the Rwanda

245 DHIS2 COVID-19 testing registry to obtain missing demographic or clinical data.¹⁸

246 Personal identifiers were removed. For the purposes of our study, a ‘conversation’ was

247 defined as a text exchange consisting of at least one ‘incoming’ patient message

248 containing three or more words (filtering out minimal replies e.g., ‘doing well’), together

249 with at least one ‘outgoing’ public health clinician or automated system message (thus

250 being interactive). Patient registration and numbers of conversations and their

251 characteristics (# messages, language used) were quantified. Conversation usage was

252 compared across patient sociodemographic and clinical characteristics (sex, age,

253 province of residence, pandemic wave, and COVID-19 status (case vs. contact)) using

254 a multivariable logistic regression model. Odds ratios (OR) are reported with 95%

255 confidence intervals (CI) and $p < 0.05$ was considered significant. Example conversations

256 were selected to illustrate context.

257

258 **Language translation and topic labelling**

259

260 A list of topics and subtopics was developed through a three-round mini-Delphi process
261 involving ten public health experts, clinicians, and researchers from Rwanda and
262 Canada. This list was an expansion of a previously published list focused on clinical
263 care and used to develop a conversation analysis and visualization tool, ConVIScope.¹²
264 The final list of topics and subtopics encompassed perspectives on public health,
265 emergency response, and clinical care. Topic definitions are provided in (S1 Fig).

266
267 For the purpose of labeling a training dataset for topic classification models, a sample of
268 2,791 text conversations from March 18, 2020, to May 31, 2021, was extracted from the
269 study period corpus. These conversations were analyzed for language use with
270 Google's CLD2 language detector. Non-English sample messages were translated into
271 English by a compensated professional healthcare text translator proficient in
272 Kinyarwanda, English, French, and Kiswahili.

273
274 The English-translated conversations were labeled by at least three trained members of
275 the research team using a custom conversation annotation tool. Any conflicts in labeling
276 were resolved by an external team member. The topic labels were then applied back to
277 the corresponding conversations in their original language, where applicable.

278

279 **Classification model training, testing, and application**

280

281 We aimed to identify the best performing classification model for each topic and
282 subtopic. Training set sample size constraints limited our classification experiments to

283 topics with ≥ 100 occurrences. We employed the binary relevance strategy, which
284 simplifies performance management by building a separate binary classifier for each
285 topic. We experimented with both traditional ML algorithms and Transformer-based
286 language models for each topic. We explored large language models (LLMs, e.g., GPT-
287 4), however, our literature review revealed that, of the three pre-trained with any
288 Kinyarwanda, the pre-training datasets for these models contained minimal
289 Kinyarwanda text ($\leq 0.01\%$). Investigations of the performance of these LLMs at NLP
290 tasks in Kinyarwanda suggested they were insufficient to include.¹⁹

291 *Traditional-ML classification models:* For training traditional ML models, we explored
292 five feature sets: binary and non-binary bag of words, TF-IDF (term-frequency inverse
293 document frequency), and character n-gram (with/without TF-IDF). The character n-
294 gram features were tried because Kinyarwanda is an agglutinative language. All
295 combinations of one of these feature sets, and one classification algorithm, were tested.
296 We evaluated logistic regression, ridge classifier, and random forest as classification
297 algorithms. Hyperparameters for each classifier were tuned using random search in
298 five-fold cross validation.²⁰

299 *Transformer-Based Language Models:* We leveraged open-source pre-trained
300 Transformer-based language models for Kinyarwanda, including AfriBERTa, AfroLM,
301 AfroXLMR, and KinyaBERT.²¹⁻²⁴ To fine-tune these models for our task, we employed
302 the AdamW optimizer along with the binary cross-entropy loss function. Additionally, we
303 incorporated dropout layers (with a probability of 0.0 for the first layer and 0.3 for
304 subsequent layers) for regularization.

305

306 We compared the performance of each model for each topic using the validation set F1-
307 score, with a minimum F1-score cut-off of 0.7. The highest F1-score models were
308 applied to predict topics for the remaining 9,328 unlabeled, untranslated SMS
309 conversations.

310
311 *Additional Language experiments:* Since our corpus contained a mix of high- and low-
312 resource languages (e.g., English, French, Kinyarwanda) we explored how this might
313 affect performance of the models by conducting identical experiments to above using
314 only the English-translated version of the training corpus. We evaluated both traditional
315 ML models (as above), and English pre-trained Transformer models (BERT,
316 Longformer).^{25,26} F1 scores are reported for comparison.

317
318 See (S2 Fig.) for more hyperparameter and fine-tuning details.

319

320 **Topic analysis and statistics**

321
322 Counts and proportions of conversations containing each topic (full corpus), and
323 subtopic (sampled corpus), are provided. Topic frequencies were compared by patient
324 demographics and clinical metrics using multiple independent multivariable logistic
325 regression models for each topic. Models incorporate all patient demographic and
326 clinical characteristics and use complete cases. OR are reported with 95% CIs and p-
327 values are adjusted for multiple testing using the Benjamini-Hochberg correction with a
328 false discovery rate of 0.05. Classified topic categories could be batch filtered and

329 manually read to understand the nature of the discussions that occurred within each
330 topic.¹² Data summarizations, visualizations, and statistical testing were produced or
331 conducted using R version 4.2.1.

332

333 Discussion

334

335 To our knowledge, this is the first study to report on the use of NLP to investigate a
336 broad range of topics from a complete corpus of conversational texts with patients in
337 home isolation in a pandemic or outbreak setting at large scale. Through a combination
338 of manual annotation and NLP techniques, we gained insights into patient experiences
339 during COVID-19 in Rwanda by identifying key discussion topics between patients and
340 public health clinicians. The sustained use of the text messaging service in Rwanda
341 demonstrated the feasibility of using SMS-based communication for remote monitoring
342 at scale in a region with limited internet access but widespread mobile phone
343 penetration. Leveraging existing infrastructure allowed for efficient support for isolated
344 individuals, regardless of location or access to healthcare facilities.

345

346 Our data captured the first two years of the Rwandan COVID-19 pandemic, including
347 text conversations in Kinyarwanda, English, and other local languages with public health
348 clinicians to report their status or seek advice. Topic classification revealed that medical
349 topics, such as symptoms, diagnostics, prevention, and treatment, were commonly
350 discussed, reflecting patient's primary focus on their health status and obtaining medical
351 guidance. However, social and lifestyle topics, including discussions about social

352 support, healthcare logistics, diet/nutrition, finances, and service quality, also emerged
353 as important aspects to patients during home isolation. The data underscores the need
354 for holistic support systems for patient isolation programs. Additionally, while our topic
355 classification approach rapidly quantified prevalent topics well, reflecting their relative
356 importance to patients, it also enables the sorting of large datasets into smaller
357 categories for easier manual evaluation by decision-makers. The identification of less
358 common topics, such as food insecurity, may also provide actionable targets for public
359 health responders, especially if it can be localized to communities.

360
361 NLP techniques are rapidly being introduced into health systems to understand
362 unstructured data from large datasets, such as clinical notes in electronic medical
363 records (EMRs), social media posts, or free-text survey responses.¹¹ For example, topic
364 classification has been employed to sort unstructured patient feedback from text
365 messages.¹⁵ Still, manual qualitative analyses dominate. A manually labeled thematic
366 analysis of 1,454 patient portal secure messaging conversations between cancer
367 patients and health professionals in the US during COVID-19 found 26% of
368 conversations were related to COVID-19 and identified topics related to changes in care
369 plans, symptoms, risks, precautions and symptoms.²⁸ In Northern Thailand manual topic
370 classification of SMS exchanges with 213 home-isolated patients with diagnosed
371 COVID-19 revealed several complexities of care including discussing symptoms or vital
372 signs, medications, diet and exercise, mental stresses, information about COVID-19,
373 and a high proportion of logistical assistance to navigate the health services.²⁹ Manual
374 labeling is resource intensive and generally limited to smaller data sets. One US-based

375 study used NLP techniques to determine if patient-initiated secure messaging via a
376 EMR portal was associated with positive testing, but this was a simple binary
377 classification.¹³ To our knowledge, our study is the first to use NLP techniques to
378 classify patient clinician texts into broad medical, social and logistical topics set by
379 experts, and at a national scale.

380
381 Our study compared performance of several ML model approaches. Traditional ML
382 models and transformer-based language models were both effective in classifying text
383 message conversations into the predefined topics. While traditional ML models
384 performed better with untranslated conversations, transformer models had superior
385 performance when trained on English (translated) data, highlighting their limited
386 performance on African languages.³⁰ This observation is consistent with the knowledge
387 that transformer architecture language models are likely to underperform on NLP tasks
388 in low-resource languages until sufficient high-quality pre-training data is sourced and
389 leveraged.³¹ Similarly, LLMs, which are also based on transformer architecture,
390 currently contain too little, if any, training on Kinyarwanda and other low-resource
391 languages to be useful in our context.¹⁹ A brief trial using Google Translate to machine
392 translate Kinyarwanda text into English for our study was inaccurate. This highlights the
393 importance of considering language-specific model performance when analyzing
394 multilingual text data.³² Language consideration remains a significant challenge, with
395 over 4,000 written languages worldwide. Pandemics affect everyone and AI tools must
396 account low-resource languages to be equitable.

397

398 Our study has limitations. First, the supervised learning approach required the pre-
399 definition and manual labeling of topics of interest which can be human resource
400 intensive. Furthermore, the topics of interest or training samples annotated may evolve
401 over time or change in different settings requiring continuously updating training
402 datasets. For example, prevalent discussions of vaccines came later in the pandemic
403 after most of our training data had been annotated. Including unsupervised ML models
404 that discover topics as they emerge could help overcome this bias, but have their own
405 flaws. Furthermore, while our dataset was similar in size to those used to train other
406 topic classification models in healthcare, many subtopics of interest were not prevalent
407 enough to develop models. Additionally, the training dataset was translated into English
408 for labelling purposes, again, being resource intensive and potentially altering meaning
409 in translation. Finally, training the models took time and conducting the NLP analyses
410 afterwards meant the data was not available to decision-makers real time. Ideally, once
411 models are trained and validated, they could be applied while an outbreak or pandemic
412 is occurring and provide insights into patient and population issues that could be
413 responded to while they are occurring. Together with location data, this type of analysis
414 could contribute to precision public health responses.

415

416 **Conclusion**

417

418 Moving forward, future research should explore ways to further optimize the evaluation
419 of text-based communication for greater insights into the patient care journey in
420 pandemics, and clinical care. It must consider technology infrastructure available to

421 populations globally and be inclusive of resource-limited languages.^{33,34} Efforts to
422 improve language support and translation capabilities across diverse linguistic and
423 cultural contexts will be essential to ensure equitable access to digital services by
424 people in all regions. Ultimately, emerging research on models that detect, understand,
425 and generate natural language, as well as development of tools that formulate
426 predictions and recommendations for healthcare decision-makers, will achieve the
427 higher accuracy, precision, and reliability required for use in routine healthcare.

428

429

430

431 **Acknowledgements**

432

433 We would like to thank the staff at the Rwanda Biomedical Center and work-learn
434 students at the University of British Columbia who helped with language translation,
435 annotation, and various aspects of background research and model building.

436

437

438 **References**

- 439 1. Reddy KP, Shebl FM, Foote JHA, Harling G, Scott JA, Panella C, et al. Cost-
440 effectiveness of public health strategies for COVID-19 epidemic control in South
441 Africa: a microsimulation modelling study. *Lancet Glob Health*. 2021
442 Feb;9(2):e120–9.
- 443 2. Mbunge E, Fashoto S, Akinnuwesi B, Gurajena C, Metfula A. Challenges of Social
444 Distancing and Self-Isolation during COVID-19 Pandemic in Africa: A Critical
445 Review [Internet]. Rochester, NY; 2020 [cited 2024 Jun 13]. Available from:
446 <https://papers.ssrn.com/abstract=3740202>
- 447 3. Wang F, Gao Y, Han Z, Yu Y, Long Z, Jiang X, et al. A systematic review and
448 meta-analysis of 90 cohort studies of social isolation, loneliness and mortality. *Nat*
449 *Hum Behav*. 2023 Aug;7(8):1307–19.
- 450 4. Babili A, Nsanzimana S, Rwagasore E, Lester RT. SMS-based digital health
451 intervention in Rwanda’s home-based care program for remote management of

- 452 COVID-19 cases and contacts: A qualitative study of sustainability and scalability.
453 Front Digit Health. 2023 Jan 9;4:1071790.
- 454 5. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a
455 mobile phone short message service on antiretroviral treatment adherence in
456 Kenya (WeITel Kenya1): a randomised trial. The Lancet. 2010 Nov
457 27;376(9755):1838–45.
- 458 6. Rathgeber SL, Hutchison SM, De Souza AM, Lester R, Blydt-Hansen T, Human
459 DG, et al. A text messaging intervention and quality of life in adolescents with solid
460 organ transplants. Pediatr Transplant. 2022;26(3):e14219.
- 461 7. van der Kop ML, Muhula S, Nagide PI, Thabane L, Gelmon L, Awiti PO, et al.
462 Effect of an interactive text-messaging service on patient retention during the first
463 year of HIV care in Kenya (WeITel Retain): an open-label, randomised parallel-
464 group study. Lancet Public Health. 2018 Mar;3(3):e143–52.
- 465 8. MAIN INDICATORS: 5th Rwanda Population and Housing Census (PHC) |
466 National Institute of Statistics Rwanda [Internet]. [cited 2024 Jun 13]. Available
467 from: https://statistics.gov.rw/publication/main_indicators_2022
- 468 9. Nachega JB, Leisegang R, Kallay O, Mills EJ, Zumla A, Lester RT. Mobile Health
469 Technology for Enhancing the COVID-19 Response in Africa: A Potential Game
470 Changer? Am J Trop Med Hyg. 2020 Jul;103(1):3–5.

- 471 10. Kop ML van der, Karanja S, Thabane L, Marra C, Chung MH, Gelmon L, et al. In-
472 Depth Analysis of Patient-Clinician Cell Phone Communication during the WeTel
473 Kenya1 Antiretroviral Adherence Trial. PLOS ONE. 2012 Sep 25;7(9):e46033.
- 474 11. Zhou B, Yang G, Shi Z, Ma S. Natural Language Processing for Smart Healthcare.
475 IEEE Rev Biomed Eng. 2024;17:4–18.
- 476 12. Li R, Hoque E, Carenini G, Lester R, Chau R. ConVIScope: Visual Analytics for
477 Exploring Patient Conversations [Internet]. arXiv; 2021 [cited 2024 Jun 13].
478 Available from: <http://arxiv.org/abs/2108.13514>
- 479 13. Mermin-Bunnell K, Zhu Y, Hornback A, Damhorst G, Walker T, Robichaux C, et al.
480 Use of Natural Language Processing of Patient-Initiated Electronic Health Record
481 Messages to Identify Patients With COVID-19 Infection. JAMA Netw Open. 2023
482 Jul 7;6(7):e2322299.
- 483 14. Lu H, Ehwerhemuepha L, Rakovski C. A comparative study on deep learning
484 models for text classification of unstructured medical notes with various levels of
485 class imbalance. BMC Med Res Methodol. 2022 Jul 2;22(1):181.
- 486 15. Cronin RM, Fabbri D, Denny JC, Rosenbloom ST, Jackson GP. A comparison of
487 rule-based and machine learning approaches for classifying patient portal
488 messages. Int J Med Inf. 2017 Sep;105:110–20.
- 489 16. Liu X, Alderman J, Laws E. A Global Health Data Divide. NEJM AI. 2024 May
490 23;1(6):Aie2400388.

- 491 17. Condo J, Uwizihiwe JP, Nsanzimana S. Learn from Rwanda's success in tackling
492 COVID-19. *Nature*. 2020 May 26;581(7809):384–384.
- 493 18. Krafft M. Going paperless for COVID-19 testing in Rwanda with DHIS2 Android
494 Capture App [Internet]. DHIS2. 2020 [cited 2024 Jun 13]. Available from:
495 <https://dhis2.org/rwanda-covid-testing/>
- 496 19. Moslem Y, Haque R, Kelleher JD, Way A. Adaptive Machine Translation with Large
497 Language Models [Internet]. arXiv; 2023 [cited 2024 Jun 13]. Available from:
498 <http://arxiv.org/abs/2301.13294>
- 499 20. Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization. *J Mach*
500 *Learn Res*. 2012;13(10):281–305.
- 501 21. Ogueji K, Zhu Y, Lin J. Small Data? No Problem! Exploring the Viability of
502 Pretrained Multilingual Language Models for Low-resourced Languages. In:
503 Ataman D, Birch A, Conneau A, Firat O, Ruder S, Sahin GG, editors. Proceedings
504 of the 1st Workshop on Multilingual Representation Learning [Internet]. Punta
505 Cana, Dominican Republic: Association for Computational Linguistics; 2021 [cited
506 2024 Jun 13]. p. 116–26. Available from: <https://aclanthology.org/2021.mrl-1.11>
- 507 22. Dossou BFP, Tonja AL, Yousuf O, Osei S, Oppong A, Shode I, et al. AfroLM: A
508 Self-Active Learning-based Multilingual Pretrained Language Model for 23 African
509 Languages [Internet]. arXiv; 2022 [cited 2024 Jun 13]. Available from:
510 <http://arxiv.org/abs/2211.03263>

- 511 23. Alabi JO, Adelani DI, Mosbach M, Klakow D. Adapting Pre-trained Language
512 Models to African Languages via Multilingual Adaptive Fine-Tuning [Internet].
513 arXiv; 2022 [cited 2024 Jun 13]. Available from: <http://arxiv.org/abs/2204.06487>
- 514 24. jean-paul/KinyaBERT-small · Hugging Face [Internet]. [cited 2024 Jun 13].
515 Available from: <https://huggingface.co/jean-paul/KinyaBERT-small>
- 516 25. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep
517 Bidirectional Transformers for Language Understanding [Internet]. arXiv; 2019
518 [cited 2024 Jun 13]. Available from: <http://arxiv.org/abs/1810.04805>
- 519 26. Beltagy I, Peters ME, Cohan A. Longformer: The Long-Document Transformer
520 [Internet]. arXiv; 2020 [cited 2024 Jun 13]. Available from:
521 <http://arxiv.org/abs/2004.05150>
- 522 27. Rwanda's literacy rate rises | National Institute of Statistics Rwanda [Internet].
523 [cited 2024 Jun 13]. Available from:
524 <https://statistics.gov.rw/publications/article/rwanda%E2%80%99s-literacy-rate-rises>
- 525 28. Alpert JM, Campbell-Salome G, Gao C, Markham MJ, Murphy M, Harle CA, et al.
526 Secure Messaging and COVID-19: A Content Analysis of Patient-Clinician
527 Communication During the Pandemic. *Telemed J E-Health Off J Am Telemed*
528 *Assoc.* 2022 Jul;28(7):1028–34.
- 529 29. Pinyopornpanish K, Nantsupawat N, Buawangpong N, Pliannuom S, Vanियapong
530 T, Jiraporncharoen W. Concerns of Home Isolating COVID-19 Patients While
531 Receiving Care via Telemedicine during the Pandemic in the Northern Thailand: A

- 532 Qualitative Study on Text Messaging. *Int J Environ Res Public Health*. 2022 May
533 28;19(11):6591.
- 534 30. Wu J, Liu X, Li M, Li W, Su Z, Lin S, et al. Clinical Text Datasets for Medical
535 Artificial Intelligence and Large Language Models — A Systematic Review. *NEJM*
536 *AI*. 2024 May 23;1(6):Alra2400012.
- 537 31. Hedderich MA, Lange L, Adel H, Strötgen J, Klakow D. A Survey on Recent
538 Approaches for Natural Language Processing in Low-Resource Scenarios
539 [Internet]. arXiv; 2021 [cited 2024 Jun 13]. Available from:
540 <http://arxiv.org/abs/2010.12309>
- 541 32. Yang LWY, Ng WY, Lei X, Tan SCY, Wang Z, Yan M, et al. Development and
542 testing of a multi-lingual Natural Language Processing-based deep learning system
543 in 10 languages for COVID-19 pandemic crisis: A multi-center study. *Front Public*
544 *Health*. 2023 Feb 13;11:1063466.
- 545 33. Adebamowo C, Callier S, Maduka OC, Akintola S, Kukucka J, Arima CG, et al.
546 Ethical Oversight of Data Science Health Research in Africa. *NEJM AI*. 2024 May
547 23;1(6):Alpc2400033.
- 548 34. Serra-Burriel M, Locher L, Vokinger KN. Development Pipeline and Geographic
549 Representation of Trials for Artificial Intelligence/Machine Learning–Enabled
550 Medical Devices (2010 to 2023). *NEJM AI*. 2023 Dec 11;1(1):Alp2300038.

551 **Supporting Information**

552 **S1 Figure.** Defined *topics and subtopics of interest were determined by a Delphi process*
 553 *contributed by expert members of the study team and incorporated into a bespoke annotation*
 554 *software used to label conversations.*

Topic	Subtopic	Definition and Examples
Symptoms	Physical	Any physical symptom, sensation, or feature (pleasant and unpleasant) that is mentioned in the conversation. Examples: pain, weight loss, burning, fatigue, shortness of breath, fever, diarrhea, muscle aches, cramping, coughing, bleeding, broken arm etc.
	Mental / Emotional	Any mental/emotional feeling, sensation, or feature (pleasant and unpleasant) that is mentioned in the conversation. Examples: confusion, depression, fear, anxiety, low energy, extreme mood, suicidal ideation, feelings of guilt, lack of interest, stigma etc.
	No Symptoms	Patient explicitly mentions not having symptoms. Examples: I have no symptoms, I feel fine.
Diagnostic Methods	Clinical	The process of identifying a disease, condition, or injury based on the signs and symptoms a patient is having and the patient's health history and physical exam.
	Laboratory/Testing	Discussion related to tests on blood, urine, or other tissues or substances taken from the body to help diagnose disease or other conditions, whether performed or not. Examples: blood tests (CBC, FBG, Hb1Ac, TSH, lipid panel, etc.), blood cultures, urinalysis, pulmonary function tests, biopsies, stool test (FIT, FOBT), etc.
	Imaging	Studies that provide a picture of the body's interior to help diagnose

		disease or other conditions. Examples: x-ray, CT, MRI, fMRI, ultrasound, PET scan, etc.
	Other	Other diagnostic methods/topic(s) not listed.
Treatment (Rx)	Medications	Pharmaceutical drug that usually requires a medical prescription to be dispensed. Examples: antibiotics, antidepressants, anti-hypertensives, amoxicillin, escitalopram, ramipril, etc. include over-the-counter medications (e.g., Tylenol)
	Procedures	Surgery, Breathing procedures, Radiation Therapy etc.
	Alternative	Non-traditional method of healing/treating disease/condition. Examples: herbal medicine, acupuncture, massage therapy, naturopathy, meditation, etc.
	Physical Therapy	Care that aims to ease pain and helps function, move, and live better. Examples: physiotherapy (geriatric, neurological, orthopedic, chiropractic etc.)
	Counselling	Giving advice/coaching/making recommendations to the patient (can be non-health related). Examples: psychologic counseling, coaching of lifestyle modifications like diet beyond standard clinical advice (e.g., in-depth discussion on how to reduce salt intake, not simply saying to reduce salt), smoking cessation, the importance of exercise on cardiovascular health, etc.
	Other	Other treatment topic(s) not listed.
Prevention	Pharmaceutical	Any mention of an intervention that utilizes chemical or biological

	Prevention	<p>methods given to or taken by humans in an attempt to prevent disease.</p> <p>Examples: prophylactic medications, immunizations, treatment as prevention, etc.</p>
	Non-Pharmaceutical Prevention	<p>Any mention of an intervention that does NOT rely on bio/chemical products given to or taken by humans in an attempt to prevent disease. Especially those methods that are behavioural (ex: physical distancing) or structural (ex: laws or policies).</p> <p>Examples: personal protective equipment (e.g., masking), physical distancing, barriers, quarantine/isolation, lockdowns, school closures, education programs, policies, laws, testing without treatment, environmental decontamination, animal/pest control, etc.</p>
Healthcare Logistics	Outpatient logistics/scheduling	<p>Mention or pertaining to setting up or attending outpatient clinical care or diagnostics.</p> <p>Examples: clinic, family doctor office, ECG appointment, hemodialysis appointment, dentist appointment, specialist appointment, physiotherapy, etc.</p>
	Hospitalization	<p>Mention or pertaining to any admission or stay in hospital.</p> <p>Receiving medical treatment during hospital admission.</p> <p>Examples: visit to emergency department, or regarding planning, admission, or discharge from hospital care. Includes long-term care facilities.</p>
Lifestyle/Behavioural	Diet / Nutrition	<p>Relating to food/liquid intake and health.</p> <p>Examples: nutrition related to weight management, apple, fruits, pie, sandwich, water, Coca-Cola, dinner, lunch, waffles, chips, popcorn,</p>

		fish, steak, etc.
	Exercise	<p>Relating to physical activity done to improve health and fitness.</p> <p>Examples: working out, going to the gym, jogging, playing soccer, playing basketball, doing yoga, walking, walking the dog, etc.</p>
	Substance Use	<p>Relating to the use of drugs or alcohol, and includes substances such as cigarettes, illegal drugs, prescription drugs, inhalants, solvents, etc.</p> <p>Examples: alcohol, marijuana, tobacco use, fentanyl, ecstasy, cocaine, addiction of substances etc.</p>
	Other	Other lifestyle topic(s) not listed.
Social/Environmental	Housing	<p>Relating to housing, shelter.</p> <p>Examples: “living at home currently,” “bought a new house,” “moving into an apartment in Victoria,” “can’t afford a place,” “got evicted,” “am homeless”.</p>
	Work/School	<p>Relating to work or school.</p> <p>Examples: “studying psychology,” “UBC,” “lots of studying to do,” “stressed from exams,” “I’m a banker,” “work was so busy,” “got fired,” etc.</p>
	Social Services	<p>Relating to government services provided for the benefit of the community, such as education, medical care, and housing programs.</p> <p>E.g.: family benefits, welfare, income assistance, housing assistance, food subsidies, foster care, disability support, mental</p>

		health services, etc.
	Friends & Family	<p>Relating to friends and family relations.</p> <p>E.g.: “went to see my mom,” “spent time with friends,” “hiking with my partner,” “family dinner,” “hung out with my best friend,”</p>
	Cultural/Religion	<p>Relating to ideas and customs of a culture.</p> <p>Examples: “Chinese herbal medicine,” “fasting for Ramadan,” “sweat lodges,” religious and spiritual practices, cultural institutions i.e., church, having own “ideas” about their own illness based on their culture, refusing conventional treatment because of what is taught in their culture.</p>
	Travel	<p>Mention or related to transportation and or travel.</p> <p>E.g., vacation in Mexico, work abroad, traveled to... I’m currently in Toronto.</p>
	Physical Environment/Climate	<p>Mention or pertaining to the climate or physical environment.</p> <p>E.g., smoke from fires, sun exposure, cold related illness, concerns re logging or farming practices</p>
	Financial	<p>Mention or pertaining to personal or other finances.</p> <p>E.g., cost of something, affordability, economics.</p>
	Other	Other social topic(s) not listed.
	Service Quality	Problems Solved
Grateful Patient		Demonstration of gratitude toward the HCP or care provided.

		Examples: Thanks! (if appears as genuine gratitude, beyond routine customary formalities). I really appreciate. (NOT: See you later, thanks!)
	Service Complaint	Conversation contains any concern or complaint brought forward about the care one is receiving, even if it also contains positive mentions.
	Request to Stop	Select this box if, at any time, the patient requests to be removed from the messaging platform or stop receiving messages.
Safety Concerns	No Subtopics	Discussing worry or concern about their life or the life of others based on specific circumstances. Example: I am worried about quarantining in a healthcentre where people are contagious.
Technical/IT	No Subtopics	Non-healthcare technical issues. Examples: trouble with texting or health IT, video conferencing and virtual care access
Maternal & Child Health	No Subtopics	Any mention or pertaining to pregnancy, breastfeeding, motherhood, and early childhood care
Health Education	No Subtopics	Providing information to help patients beyond simple advice, with the intent to increase knowledge - can be verbal, written or in the form of visual aids like pamphlets, courses. Examples: pamphlet on what diabetes is, teaching how to self-monitor blood pressure, etc.
Stigma	No Subtopics	A mention of the disapproval of, or <u>discrimination</u> against, an individual or group based on perceivable characteristics that serve to distinguish them from other members of a society. Social stigmas are commonly related to <u>culture</u> , <u>gender</u> , <u>race</u> , <u>socioeconomic class</u> , age, sexual orientation, <u>body image</u> , intelligence or lack

thereof, and health. Can be self-directed.

555

556

557 **S2 Figure. NLP Model Details**

<i>Traditional machine learning feature extraction details</i>					
Feature Extractor		Detail			
Character n-gram		Experimented with word bound character ranges of 1-3, 1-4, and 1-5.			
<i>Traditional machine learning model hyperparameters and random search ranges</i>					
Model		Hyperparameter		Range	
Logistic Regression		C (inverse of regularization strength)		Loguniform (1e-4, 1e4)	
		Penalty		L1, L2	
Random Forest		Max_features (number of features to consider when looking for best split)		Uniform (0,1)	
		Min_samples_leaf (minimum number of samples required to be at a leaf node)		Randint (1, 20)	
Ridge Classifier		Alpha (regularization strength)		Loguniform (1e-2, 1e3)	
<i>Transformer model hyperparameters</i>					
Model	Epochs	Learning Rate	Loss	Optimizer	Batch Size
AfriBERT ^a	15	Starts at 0.01*	BCEWithLogitsLoss	AdamW	32
AfroXLMR ^b	15				
AfroLM ^c	15				
KinyaBERT ^d	15	10 ⁻⁵			

558 * We utilized the ReduceLROnPlateau scheduler that reduces the learning rate by a factor of 0.5 for every
 559 two epochs where the validation loss does not decrease, up to a minimum value of 10⁻⁵.

560 ^a <https://aclanthology.org/2021.mrl-1.11/>

561 ^b <https://arxiv.org/abs/2204.06487>

562 ^c <https://arxiv.org/abs/2211.03263>

563 ^d <https://huggingface.co/jean-paul/KinyaBERT-small>

564

565

566 **S3 Figure.** *Frequency and proportion of conversations in which all topics and subtopics of*

567 *interest occurred within the human-labelled, machine-labelled, and complete corpora. Note:*

568 *Conversations were often labelled with multiple topics/subtopics, therefore percentages do not*

569 *sum to 100%.*

	Human Labelled (n=2,791)	Machine Labelled (n=9,328)	Human-and-Machine-Labelled (n=12,119)
Topic Subtopic(s)	n (%)	n (%)	n (%)
Symptoms	2,002 (72)	6,305 (68)	8,307 (69)
None	1,657 (59)	-	-
Physical	406 (15)	-	-
Mental/Emotional	45 (1.6)	-	-
Diagnostic Methods	1,030 (37)	3,539 (38)	4,569 (38)
Laboratory/Testing	1,018 (36)	-	-
Imaging	7 (0.2)	-	-
Clinical	5 (0.2)	-	-
Other	5 (0.2)	-	-
Prevention	459 (16)	1,685 (18)	2,114 (18)
Pharmaceutical	17 (0.6)	-	-
Non-pharmaceutical	443 (16)	-	-

Service Quality	425 (15)	NP	NP
Grateful Patient	244 (8-7)	-	-
Service Complaint	106 (3-8)	-	-
Request to Stop	50 (1-8)	-	-
Problems Solved	45 (1-6)	-	-
Social	406 (15)	1,896 (20)	2,302 (19)
Cultural/Religion	154 (5-5)	-	-
Friends & Family	138 (4-9)	-	-
Financial	53 (1-9)	-	-
Work/School	46 (1-7)	-	-
Housing	32 (1-2)	-	-
Social Services	30 (1-1)	-	-
Travel	12 (0-4)	-	-
Environment/Climate	6 (0-2)	-	-
Other	7 (0-3)	-	-
Healthcare Logistics	281 (10)	1,599 (17)	1,880 (16)
Logistics/Scheduling	268 (9-6)	-	-
Hospitalization	15 (0-5)	-	-
Technical/IT	185 (6-6)	NP	NP
Treatment	174 (6-2)	791 (8-5)	965 (8-0)
Medications	152 (5-5)	-	-
Alternative	35 (1-2)	-	-

Physical Therapy	1 (0.04)	-	-
Procedures	0 (0)	-	-
Counseling	0 (0)	-	-
Other	0 (0)	-	-
Lifestyle/Behavioural	156 (5.5)	NP	NP
Diet/Nutrition	102 (3.7)	-	-
Exercise	25 (0.9)	-	-
Substance Use	0 (0)	-	-
Other	52 (1.9)	-	-
Health Education	45 (1.6)	NT	NT
Maternal & Child Health	33 (1.2)	NT	NT
Safety Concern	20 (0.7)	NT	NT
Stigma	15 (0.5)	NT	NT

570 NP = Not Predictable (best model did not meet F1 performance cutoff and was not applied)

571 NT = Not Tested (insufficient data to develop models)

572

573

574 **S4 Table.** Comparison of the odds of discussion of predictable conversation topics of interest
 575 (i.e., those conversation topics of interest for which we developed a classification model of
 576 interest meeting the $F1 \geq 0.7$ performance cutoff) by patient demographic and clinical factors in
 577 the complete conversation corpus ($n=12,119$). Presented p-values were adjusted for multiple
 578 testing using the Benjamini-Hochberg correction.

TOPIC	COMPARISON	ODDS RATIO (95% CI, adj. p-value)
-------	------------	-----------------------------------

	(vs. reference)	
Symptoms	Female vs. Male	1.04 (0.96-1.13, p=0.39)
	Age (per year increase)	1.00 (1.00-1.00, p=0.47)
	Contact vs. Case	1.25 (1.06-1.47, p=0.02)
Diagnostic Methods	Female vs. Male	0.87 (0.80-0.94, p=0.0012)
	Age (per year increase)	1.00 (0.99-1.00, p=0.060)
	Contact vs. Case	0.84 (0.73-1.00, p=0.060)
Social	Female vs. Male	1.04 (0.94-1.14, p=0.45)
	Age (per year increase)	1.005 (1.002-1.009, p=0.012)
	Contact vs. Case	1.09 (0.91-1.30, p=0.43)
Prevention	Female vs. Male	0.93 (0.84-1.02, p=0.29)
	Age (per year increase)	1.00 (0.99-1.00, p=0.22)
	Contact vs. Case	0.61 (0.49-0.75, p=0.00004)
Healthcare Logistics	Female vs. Male	0.88 (0.79-0.98, p=0.025)
	Age (per year increase)	1.005 (1.0009-1.008, p=0.025)
	Contact vs. Case	0.69 (0.55-0.86, p=0.003)
Treatment	Female vs. Male	1.077 (0.94,1.23, p=0.36)
	Age (per year increase)	1.006 (1.0009-1.011, p=0.048)
	Contact vs. Case	0.30 (0.20-0.46, p<0.0001)

579

580

