Use of Artificial Intelligence to Identify and Support Care partners in Patient Portal Messages

Kelly Gleason, PhD, RN,[a] Athena DeGennaro, BS,[a] Danielle Powell, PhD,[b] Mingche MJ Wu, MS,[b] Talan Zhang, MS,[b] Lum, Hillary, MD, PhD, Portz,[c] Jennifer, PhD,[c] Jennifer Wolff, PhD,[b]

a. Johns Hopkins University School of Nursing, Baltimore, MD; b. Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; c. University of Colorado School of Medicine

Background

We sought to assess the feasibility of identifying secure messages from nonpatient authors using a 1) a rule-based natural language processing (NLP) algorithm, and 2) common large language models, including Gemini, Meta AI, and ChatGPT.

Methods

We conducted a two-part observational study. First, two independent reviewers manually coded a randomly selected sample of 1,973 portal messages to create a gold standard for nonpatient author identification. We then developed and tested a natural language processing (NLP) rule-based classification algorithm. We compared its performance to three common large language models, Gemini, Meta AI, and ChatGPT.

Results

A total of 267 (13.53%) of 1,973 messages sent from older adults' portal accounts were identified through manual coding as sent by a nonpatient author. The rule-based NLP algorithm performance to identify nonpatient authors demonstrated an AUC of 0.90. The large language models each demonstrated a higher AUC: ChatGPT (0.92), Meta AI (0.97), and Gemini (0.99). In both methods, the messages which were least likely to be correctly ident identified were brief with limited context, or the person referenced both themselves and their partner.

Acknowledgements

This work was supported by NIH NIA grant R35AG072310, Consumer Health Information Technology to Engage and Support ADRD Caregivers: Research Program to Address ADRD Implementation Milestone 13.I, and the Hopkins' AITC, National Institute on Aging grant P30AG073104.

Classification using ChatGPT, Meta AI, and Gemini

The prompt was intentionally structured to assume that the large language model had no other information about the patient's identity, to optimize privacy and simplicity if it were to be operationalized:

Based on the provided message, guess about whether it was likely sent by the patient themselves or by someone else using the patient's account. Please consider only the content of the message.

Large language model	Area Under the Curve
ChatGPT	0.92
Meta AI	0.97
Gemini	0.99

Each large language model had difficulty with messages ones where people referenced both themselves and their partners, e.g.: "*My husband and I both received our first vaccine shots yesterday.*"

Rule-Based Natural Language Processing Algorithm

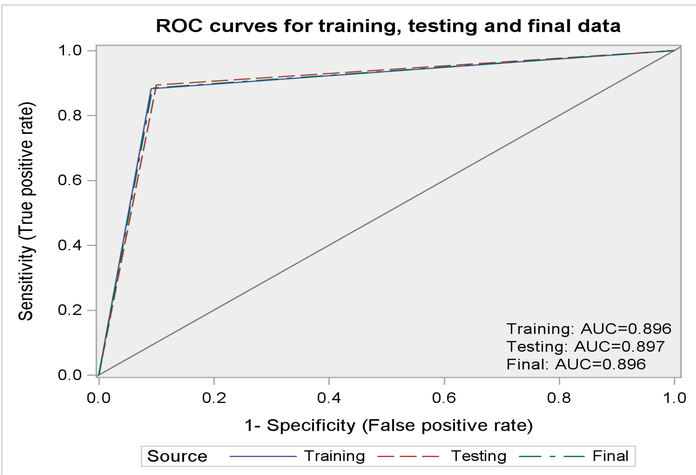
```
Text_df <- msg %>%
  mutate(Email_Text= gsub(x = Email_Text, pattern = "[0-9]", replacement = "")) %>%
  unnest_tokens(input=Email_Text, output=word)%>%
  filter(!word %in% c("message","questionnaire","dr","md",
                      "patient","hi","hello","dear","pm","subject","morning","afternoon",
                      "insert location","mg","tablet","est","edt","insert institution","mins"))


Text_df$child<-
regexpr("\\b(mom|moms|mommy|dad|dads|daddy|daddys|mother|mothers|father|fath
ers|stepmother|stepmothers|stepfather|stepfathers|stepmom|stepmoms|stepdad|stepd
ads)\\b",Text_df$word,ignore.case = TRUE)

Text_df$wife <- regexpr("\\b(wife|bwifes|wifee)\\b",Text_df$word,ignore.case = TRUE)

Text_df$husband<- regexpr("\\b(husband|husbands)\\b",Text_df$word,ignore.case =
TRUE)

Text_df$partner<- regexpr("\\b(partner|spouse)\\b",Text_df$word,ignore.case = TRUE)

Text_df$grandchild<-
regexpr("\\b(grandma|grandmas|grandpa|grandpas|grandmother|grandmothers|grandfat
her|grandfathers|grandparent|grandparents)\\b",Text_df$word,ignore.case = TRUE)

Text_df$thirdpron<-
regexpr("\\b(his|him|he|she|hers|her)\\b",Text_df$word,ignore.case = TRUE)
```

Similar to the design of the prompt, the algorithm was intentionally created to optimize for privacy and simplicity.

Sample Characteristics

	Patients (n=1,973)
Age	74.20
Female Sex	1,091, 55.30%
Race	
White	1,455, 74.92%
Black	353, 18.18%
Asian	45, 2.32%
Hispanic Ethnicity	29, 1.49%
Dementia diagnosis	108, 5.47%
Shared Access (Proxy) registered	124, 6.28%
Patient portal metric	28.95

Precision of the Model



Conclusion

A rule-based natural language processing algorithm performed similarly to commonly used large language models. Thus, there is not a substantial tradeoff in precision when deciding which to use.