Voice Privacy Assistant for Monitoring In-home Voice Commands

Bang Tran[1], Xiaohui Liang[1], Gabriel Ghinita[2], Caroline Summerour[3], and John A. Batsis[3]

[1]Department of Computer Science, University of Massachusetts in Boston, MA, USA

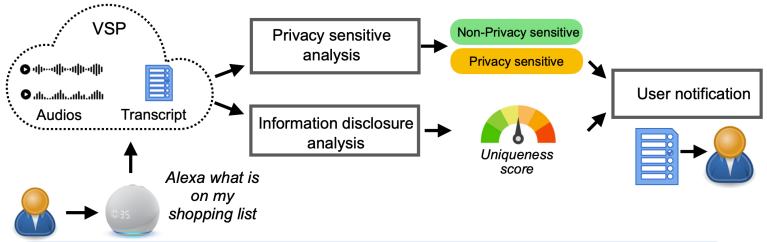
[2]College of Science & Engineering, Hamad Bin Khalifa University, Doha, Qatar

[3]Division of Geriatric Medicine, School of Medicine, University of North Carolina at Chapel Hill, NC, USA

Grant#: NIH-R01AG067416
NSF-IIS-1909806

1. Introduction

- Voice assistant systems (VAS) provide convenient means for users to interact verbally with online services and control smart home devices.
- Voice commands contain highly-sensitive information about individuals, and sharing such data with service providers must be done in a carefully controlled and transparent manner in order to prevent privacy breaches.



2. VPASS framework

- VPASS analyzes the information disclosure of in-home voice commands
- VPASS analyzes the privacy sensitivity of voice command
- VPASS generates monthly reports and immediately alert

3. Participants and annotators

- 15 old adults participants (≥65 years).
- Each older has one or more Alexa Echo devices installed at home
- Five annotators assign sensitive label for their commands

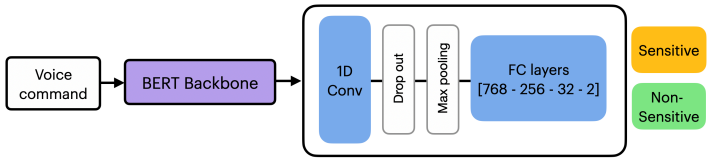
Semantical-similarity of two commands c_i and c_j

$$s_{i,j} = \text{Sim}(c_i, c_j) = \frac{v_i \cdot v_j}{\| v_i \| \cdot \| v_j \|}$$

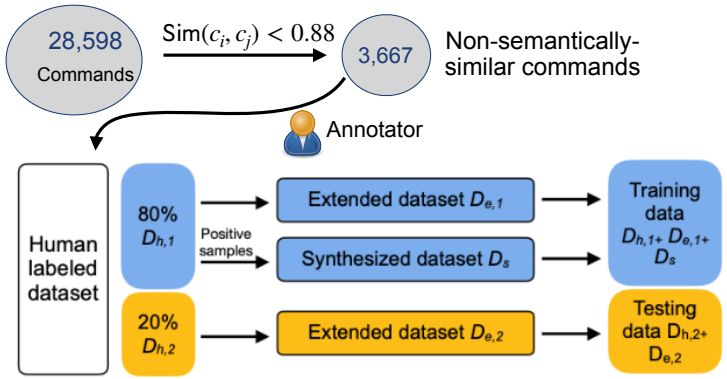
Uniqueness score of command c_{n+1}

$$\text{Uniq}(c_{n+1}, \mathbf{C}_n) = 1 - \max_{c_j \in \mathbf{C}_n} s_{n+1,j}$$

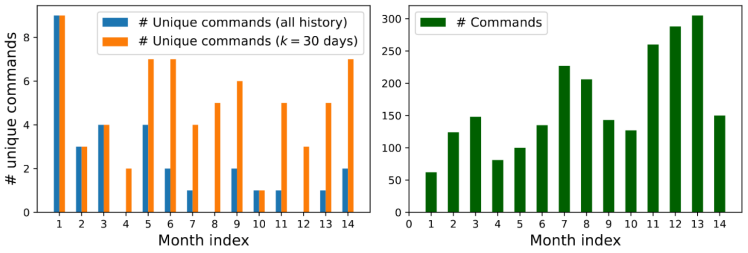
4.1. Sensitivity inference model



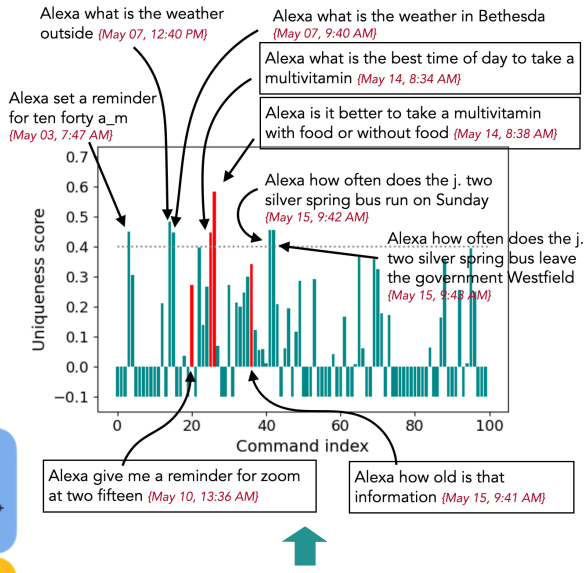
4.2. Labeling, training, and testing



5.1. The number of unique commands of User 008 (*example*)



5.2. VPASS monthly report of User 008



User 008's VPASS Report in May 2022

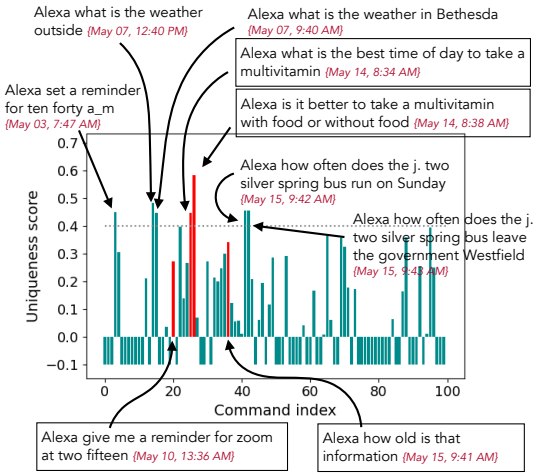
- The red commands are sensitive, detected by our sensitive inference model
- Personalized privacy alert policy
  - Seven alerts if the user only uses the uniqueness score $\geq th = 0.4$
  - Four alerts if the commands are determined as sensitive
  - Two alerts if adopting the above two conditions

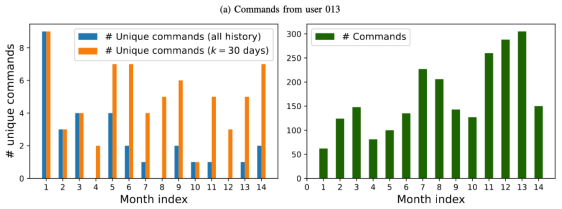
VPASS provides an intuitive interface to manage privacy requirement of VAS

* We thank Professor David Kotz (Dartmouth College) and Professor Brian MacWhinney (Carnegie Mellon University) for their feedback about this work.

- Voice assistant systems (VAS) provide convenient means for users to interact verbally with online services and control smart home devices.

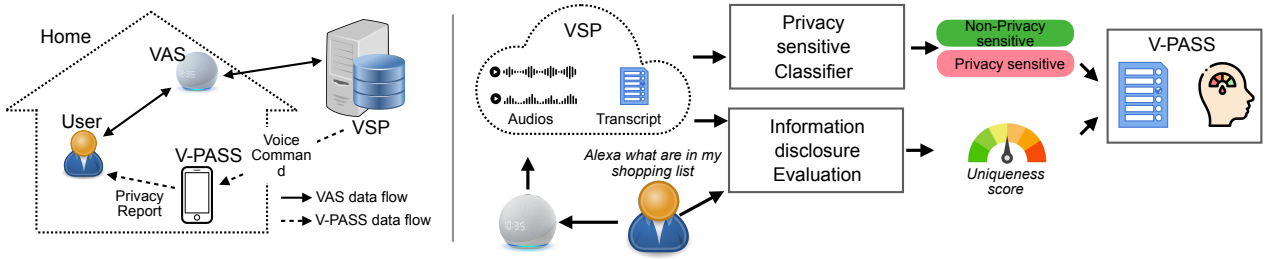
- Voice commands contain highly-sensitive information about individuals, and sharing such data with service providers must be done in a carefully controlled and transparent manner in order to prevent privacy breaches.

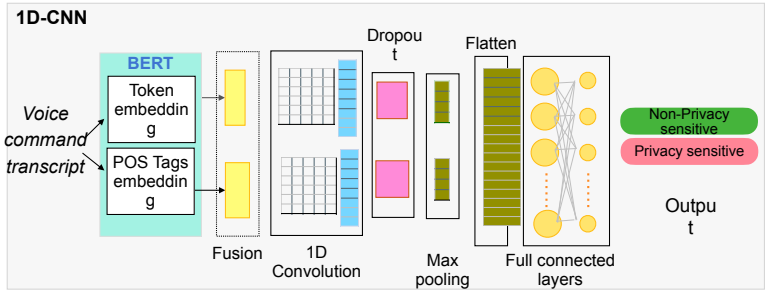


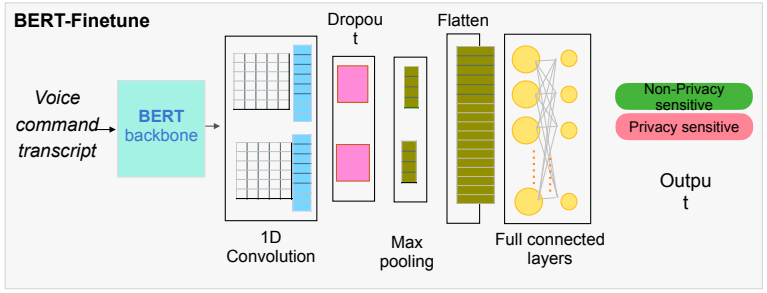


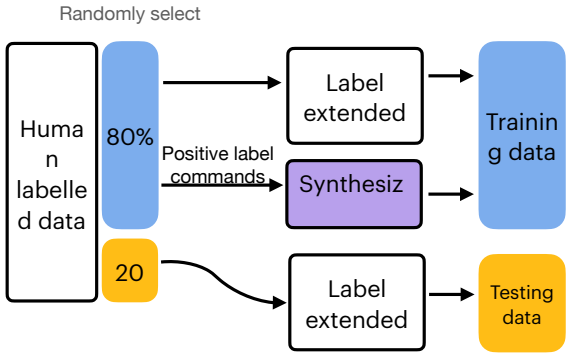
Voice command uniqueness scores

i	c_i	max(sim)	uniqueness
1	Alexa play George gGershwin music	-	1.0
2	Alexa play Leonard Bernstein music	$\{c_1\}$	0.2100
3	Alexa play Carole King music	$\{c_1\}$	0.3132
4	Alexa what is the weather in Friendship Heights	$\{c_3\}$	0.5154
5	Alexa play classical music and turn it off	$\{c_3\}$	0.3772
6	Alexa play classical music	$\{c_3\}$	0.2613
7	Alexa level four	$\{c_3\}$	0.5265
8	Alexa level four	$\{c_7\}$	0.0
9	Alexa play Hawaiian music	$\{c_8\}$	0.2901
10	Alexa play George Gershwin music	$\{c_1\}$	0.0



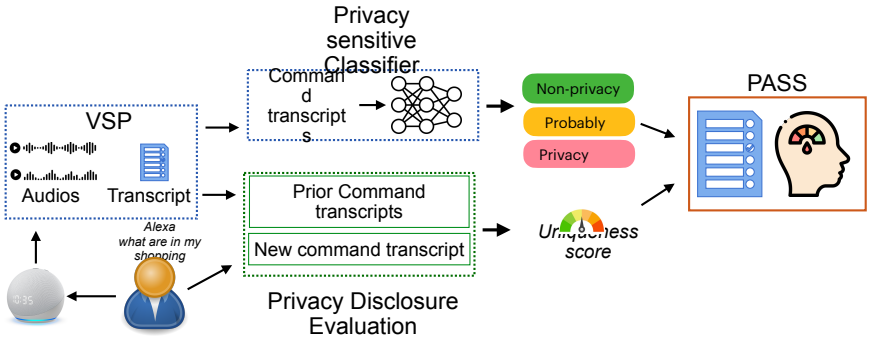


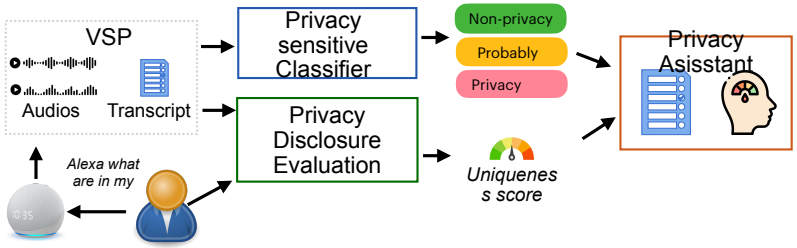


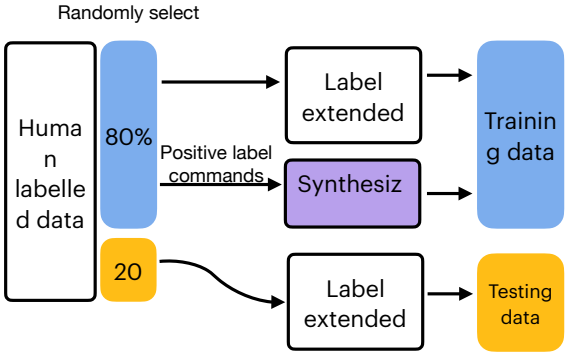


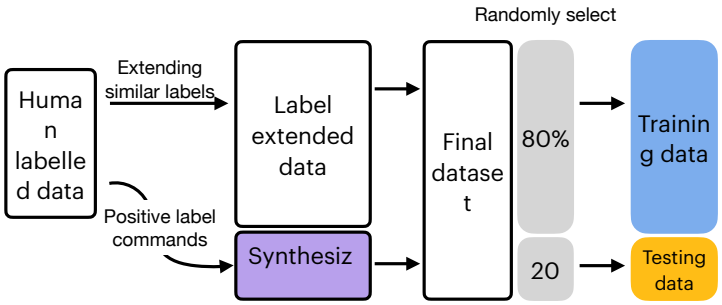
Sensitiv

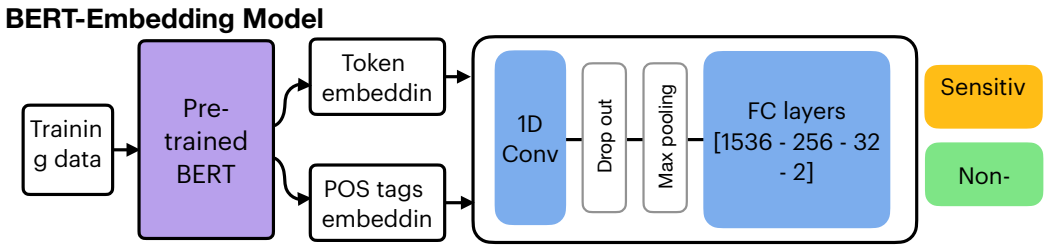
Non-

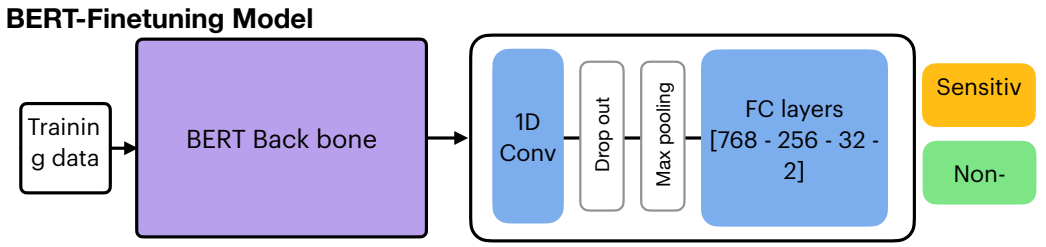

Privacy sensitive Classifier
VSP
Audios
Transcript
Command transcripts
Non-privacy
Probably
Privacy
PASS
Prior Command transcripts
New command transcript
Uniqueness score
Alexa what are in my shopping
Privacy Disclosure Evaluation

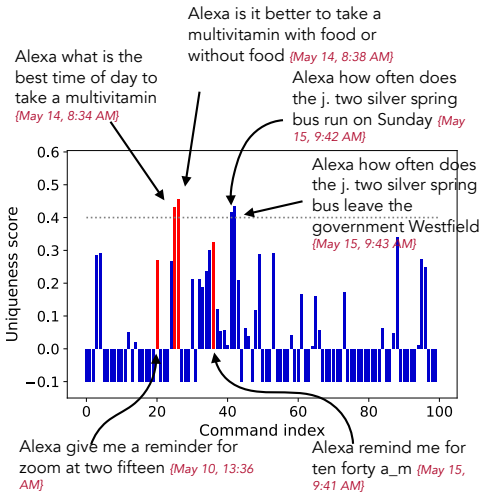

VSP
Audios
Transcript
Privacy sensitive Classifier
Privacy Disclosure Evaluation
Non-privacy
Probably
Privacy
Uniquenes s score
Privacy Asisstant
Alexa what are in my

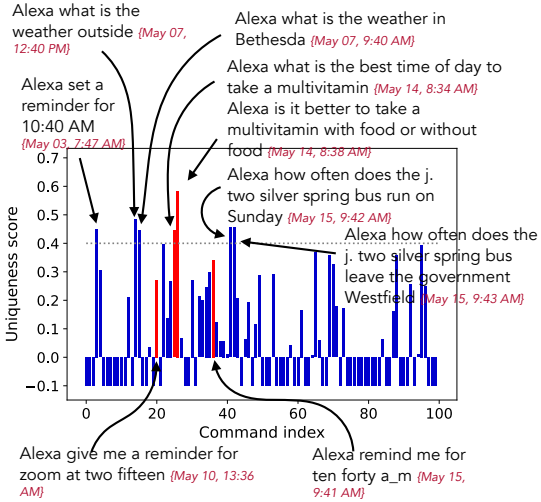

Randomly select
Huma n labelle d data
80%
20
Positive label commands
Label extended
Synthesiz
Label extended
Trainin g data
Testing data

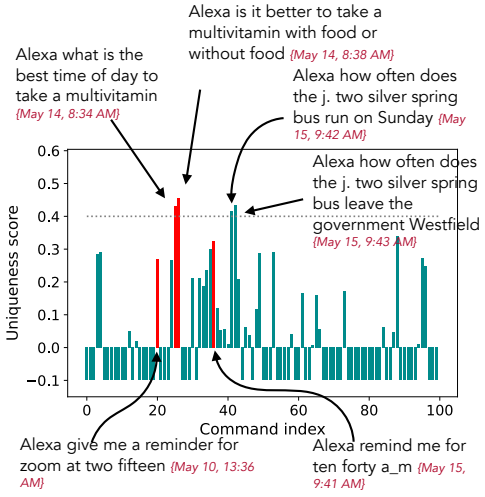

Randomly select
Huma n labelle d data
Extending similar labels
Positive label commands
Label extended data
Synthesiz
Final datase t
80%
20
Trainin g data
Testing data

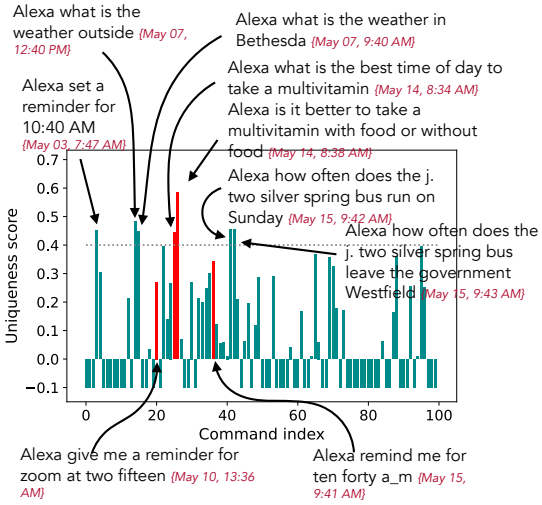

BERT-Embedding Model
Trainin g data
Pre-trained BERT
Token embeddin
POS tags embeddin
1D Conv
Drop out
Max pooling
FC layers [1536 - 256 - 32 - 2]
Sensitiv
Non-


BERT-Finetuning Model
Trainin g data
BERT Back bone
1D Conv
Drop out
Max pooling
FC layers [768 - 256 - 32 - 2]
Sensitiv
Non-

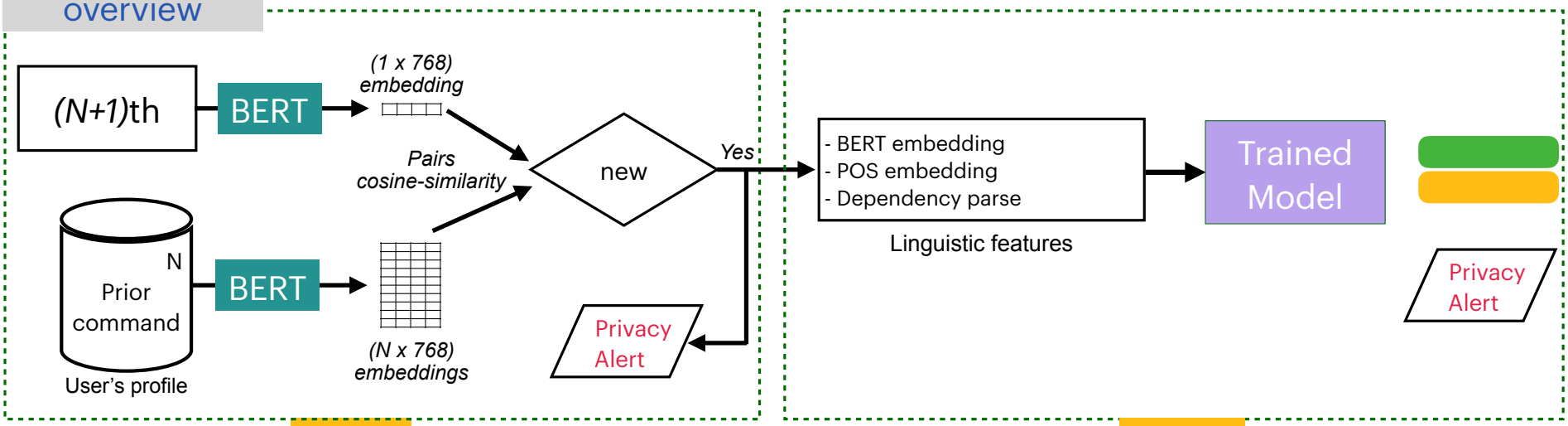

Alexa what is the best time of day to take a multivitamin {May 14, 8:34 AM}
Alexa is it better to take a multivitamin with food or without food {May 14, 8:38 AM}
Alexa how often does the j. two silver spring bus run on Sunday {May 15, 9:42 AM}
Alexa how often does the j. two silver spring bus leave the government Westfield {May 15, 9:43 AM}
Uniqueness score
Command index
Alexa give me a reminder for zoom at two fifteen {May 10, 13:36 AM}
Alexa remind me for ten forty a_m {May 15, 9:41 AM}


Alexa what is the weather outside {May 07, 12:40 PM}
Alexa what is the weather in Bethesda {May 07, 9:40 AM}
Alexa what is the best time of day to take a multivitamin {May 14, 8:34 AM}
Alexa set a reminder for 10:40 AM {May 03, 7:47 AM}
Alexa is it better to take a multivitamin with food or without food {May 14, 8:38 AM}
Alexa how often does the j. two silver spring bus run on Sunday {May 15, 9:42 AM}
Alexa how often does the j. two silver spring bus leave the government Westfield {May 15, 9:43 AM}
Uniqueness score
Command index
Alexa give me a reminder for zoom at two fifteen {May 10, 13:36 AM}
Alexa remind me for ten forty a_m {May 15, 9:41 AM}


Alexa what is the best time of day to take a multivitamin {May 14, 8:34 AM}
Alexa is it better to take a multivitamin with food or without food {May 14, 8:38 AM}
Alexa how often does the j. two silver spring bus run on Sunday {May 15, 9:42 AM}
Alexa how often does the j. two silver spring bus leave the government Westfield {May 15, 9:43 AM}
Uniqueness score
Command index
Alexa give me a reminder for zoom at two fifteen {May 10, 13:36 AM}
Alexa remind me for ten forty a_m {May 15, 9:41 AM}


Alexa what is the weather outside {May 07, 12:40 PM}
Alexa what is the weather in Bethesda {May 07, 9:40 AM}
Alexa what is the best time of day to take a multivitamin {May 14, 8:34 AM}
Alexa set a reminder for 10:40 AM {May 03, 7:47 AM}
Alexa is it better to take a multivitamin with food or without food {May 14, 8:38 AM}
Alexa how often does the j. two silver spring bus run on Sunday {May 15, 9:42 AM}
Alexa how often does the j. two silver spring bus leave the government Westfield {May 15, 9:43 AM}
Uniqueness score
Command index
Alexa give me a reminder for zoom at two fifteen {May 10, 13:36 AM}
Alexa remind me for ten forty a_m {May 15, 9:41 AM}

System overview

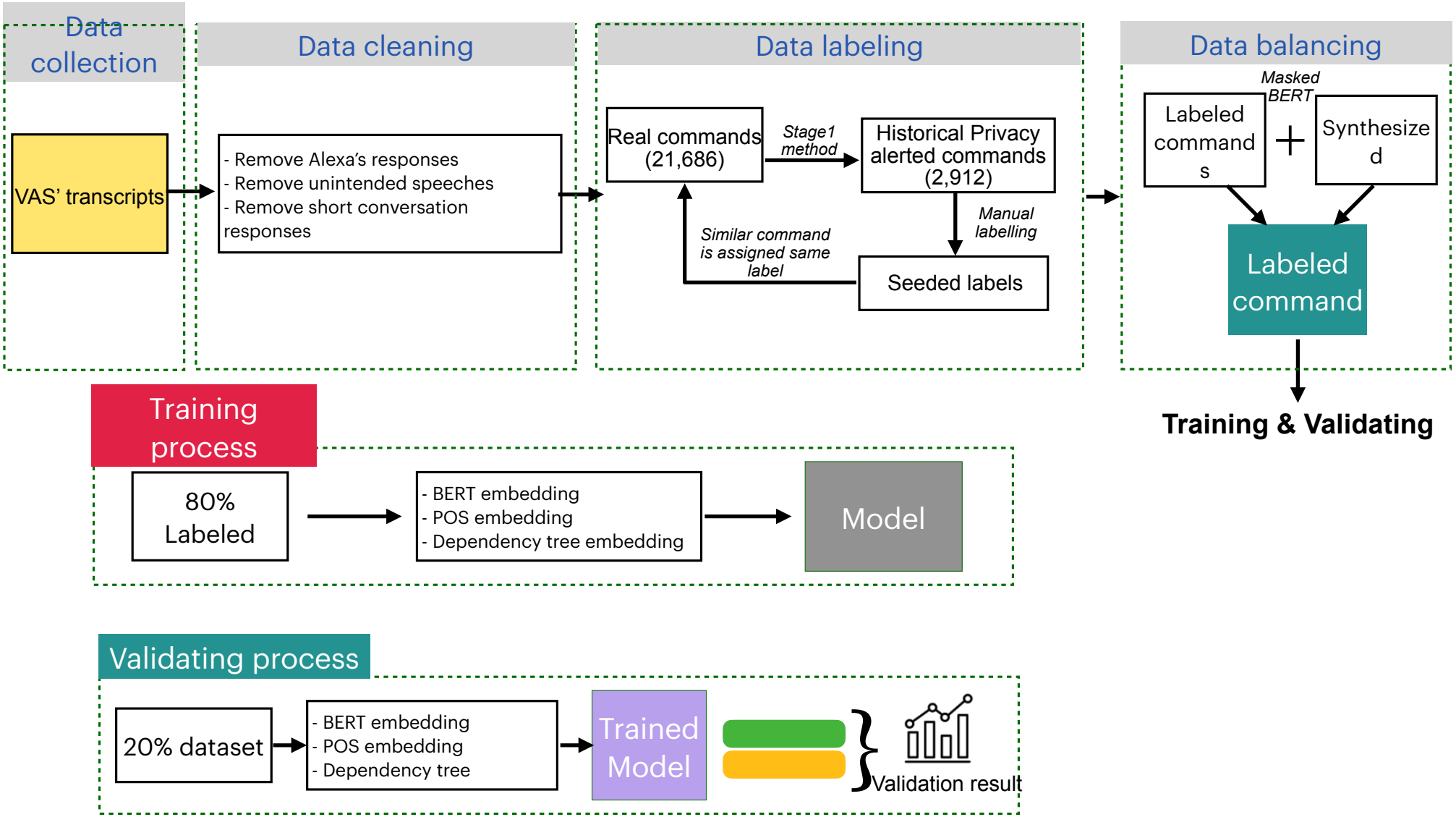


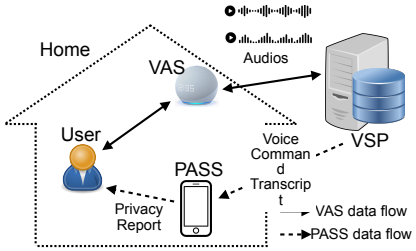
Stage 1

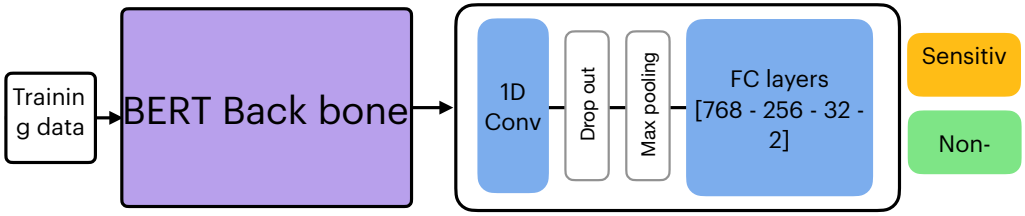
- Determines new command
- Reduces labeling efforts
- Determines real user's command templates for labels balancing process

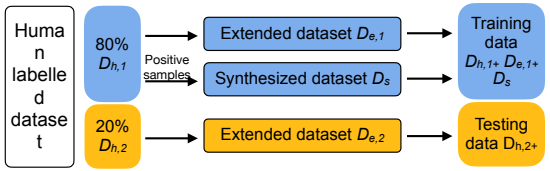
Stage 2

- Determines semantic privacy commands
- <place holder>
- <place holder>


Data collection
VAS' transcripts
Data cleaning
- Remove Alexa's responses
- Remove unintended speeches
- Remove short conversation responses
Data labeling
Real commands (21,686)
Stage1 method
Historical Privacy alerted commands (2,912)
Manual labelling
Seeded labels
Similar command is assigned same label
Data balancing
Masked BERT
Labeled commands
Synthesized
Labeled command
Training & Validating
Training process
80% Labeled
- BERT embedding
- POS embedding
- Dependency tree embedding
Model
Validating process
20% dataset
- BERT embedding
- POS embedding
- Dependency tree
Trained Model
Validation result


Home
VAS
Audios
User
VSP
Voice
Command
Transcript
PASS
Privacy
Report
VAS data flow
PASS data flow


Training data
BERT Back bone
1D Conv
Drop out
Max pooling
FC layers [768 - 256 - 32 - 2]
Sensitiv
Non-


Human labelled dataset
80% $D_{h,1}$
20% $D_{h,2}$
Positive samples
Extended dataset $D_{e,1}$
Synthesized dataset D_s
Extended dataset $D_{e,2}$
Training data $D_{h,1+}$ $D_{e,1+}$ D_s
Testing data $D_{h,2+}$