NATIONAL SECURITY AGENCY
CENTRAL SECURITY SERVICE

(U) Classification Guide for
Human Language Technology (HLT) Models
2-20

Effective Date: 18 May 2011

CLASSIFIED BY: ▉▉▉▉▉▉▉▉
       Deputy Director for Analysis
       and Production

Classification Category: 1.4 (c)

Declassify On: 25 years*

ENDORSED BY: ▉▉▉▉▉▉▉▉
       Deputy Associate Director for
       Policy and Records

CLASSIFICATION GUIDE TITLE/NUMBER:
(U) Human Language Technology (HLT) Models, 2-20

PUBLICATION DATE: 18 May 2011

OFFICE OF ORIGIN: (U) R66E, Human Language Technology Research

POC: (U//FOUO) ███████ R66E; 961-3032s

ORIGINAL CLASSIFICATION AUTHORITY: (U//FOUO) ███████ Deputy Director for Analysis and Production

Description of Information	Classification/ Markings	Category	Declass	Remarks
A. (U) General				
A.1. (U) The fact that NSA/CSS has created HLT models used for: • Gender recognition • Language • Language variety/dialect recognition • Speaker recognition • Speech-to-text processing • Speech activity detection • Anomaly detection • Phonetic recognition	UNCLASSIFIED	N/A	N/A	
A.2. (U) The fact that HLT models are obtained, at least in part, by aggregating statistics derived from SIGINT collection	UNCLASSIFIED	N/A	N/A	
A.3. (U) The fact that HLT models allow for collected audio files to be sorted and prioritized for linguists	UNCLASSIFIED	N/A	N/A	
A.4. (U) The fact that statistics in a model can be generated from one or many audio files	UNCLASSIFIED	N/A	N/A	
A.5. (U) The fact that new models are regularly generated, adding to the aggregate nature of the model	UNCLASSIFIED	N/A	N/A	
A.6. (U) The fact that SIGINT voice collection (not further identified) can be identified as: • male or female • a specific language • a specific language variety/dialect • a specific speaker • a sequence of words • speech or nonspeech	UNCLASSIFIED	N/A	N/A	(U) Further details such as which specific language, or dialect, or speaker are classified. Consult applicable SIGINT guidance.
A.7. (U) HLT models used for: • Gender recognition • Language recognition	See Remarks			(U//FOUO) The classification of HLT models used for Gender and Language Recognition is dependent upon the classification of the messages used to train the model, up to SECRET//REL TO

				USA, AUS, CAN, GBR, NZL. Although it is possible that the messages used to train the model may have a higher classification and/or more restrictive releasability than SECRET//REL, the original audio cannot be recovered from the model. SECRET//REL is sufficient to protect this type of model. (U) The Deputy Director for Analysis and Production may approve, on a case-by-case basis, foreign release of models containing otherwise non-releasable information.
A.8. (U) HLT speaker recognition models	See Remarks.			(U) Consult applicable SIGINT guidance: Classification and foreign releasability should be in accordance with the highest classification and most restrictive releasability that applies to the targeted entities used in the model. (U) The Deputy Director for Analysis and Production may approve, on a case-by-case basis, foreign release of models containing otherwise non-releasable information.
A.9. (U) HLT acoustic models used for • Speech-to-text • Phonetic tokenization	See Remarks.			(U//FOUO) The classification of HLT acoustic models is dependent upon the classification of the messages used to train the model, up to SECRET//REL TO USA, AUS, CAN, GBR, NZL. Although it is possible that the messages used to train the model may have a higher classification and/or more restrictive releasability than SECRET//REL, the original audio cannot be recovered from the model. SECRET//REL is sufficient to protect this type of model. (U) The Deputy Director for Analysis and Production may approve, on a case-by-case basis, foreign release of models containing otherwise non-releasable information.
A.10. (U) HLT language models used for • Speech-to-text	See Remarks.			(U) Consult applicable SIGINT guidance: Classification and foreign releasability should be in

• Phonetic tokenization				accordance with the highest classification and most restrictive releasability that applies to the targeted entities and/or content of the messages used in the model. (U) The Deputy Director for Analysis and Production may approve, on a case-by-case basis, foreign release of models containing otherwise non-releasable information.
A.11. (U) Speech activity detection models using syllable rate speech activity detection (SRSAD)	UNCLASSIFIED// FOR OFFICIAL USE ONLY	N/A	N/A	
A.12. (U) Anomaly detection models	UNCLASSIFIED// FOR OFFICIAL USE ONLY	N/A	N/A	
B. (U) Model Output				
B.1 (U) Output of language recognition models	UNCLASSIFIED// FOR OFFICIAL USE ONLY	N/A	N/A	(U) Results generally indicate the recognized language and the degree of confidence in the determination, e.g. "Farsi with 90% confidence." This information may require protection as classified when combined with other details regarding the input data.
B.2. (U) Output of gender recognition models	UNCLASSIFIED// FOR OFFICIAL USE ONLY	N/A	N/A	(U) Results generally indicate the recognized gender and the degree of confidence in the determination, e.g. "Male with 75% confidence." This information may require protection as classified when combined with other details regarding the input data.
B.3. (U) Output of speaker recognition models	See Remarks.			(U) Classification and foreign releasability of the results should be the same as the input data.
B.4. (U) Output of acoustic speech-to-text and phonetic tokenization models	See Remarks.			(U) Classification and foreign releasability of the results should be the same as the input data.
B.5. (U) Output of language speech-to-text and phonetic tokenization models	See Remarks.			(U) Classification and foreign releasability of the results should be the same as the input data unless the results reveal specific information used in the model that is protected at a higher level than the input data; in this case, the results require protection at the level of the model.

(U) Note: Declassification in 25 years indicates that the information is classified for 25 years from the date a document is created or 25 years from the date of this original classification decision, whichever is later.