SLP Final Examination Name _____

December 14, 2023 (sanitized)

Closed Book, Open Notes (three handwritten pages), 165 minutes.

There are no in-and-out privileges. Please answer concisely.

1. [15] True or False

   a) **T F** If Model A has higher accuracy than Model B for detecting is-depressed from voice, then A will also have higher accuracy than B for detecting not-depressed.
   b) **T F** A model that performs well on the test data will probably perform as well or better on the training data.
   c) **T F** In general, if Model A has higher precision than Model B on some data, its recall will also be higher.
   d) **T F** Self-supervised learning requires structured data for training.
   e) **T F** Named Entities in English are, syntactically, verb phases or prepositional phrases.
   f) **T F** Speech Recognition is the problem of mapping from the acoustic signal to the inferred positions of the articulators (tongue, lips, etc.).
   g) **T F** "Training a model," means giving it data from which it can learn the structure of the solution, for example how many neural layers, how many neurons per layer, which layers are transformers, and so on.
   h) **T F** Bayes Law only holds you have no information on the prior probabilities.
   i) **T F** English has 26 phonemes.
   j) **T F** English has 5 vowels.
   k) **T F** The International Phonetic Alphabet (IPA) has a symbol for every sound that exists in any of the world's languages.
   l) **T F** For diagnosing cognitive and thought disorders, both language-related features and speech-related features are informative, depending on the disease.
   m) **T F** Large language models internally use word embeddings.
   n) **T F** Pretraining is always necessary before training.
   o) **T F** Modern search engines, such as Bing, no longer use the vector-space model.
   p) **T F** Diarization is the task of determining who's speaking when in a multi-speaker recording.

2. [1] Pretrained models are useful for _____ .
{classification tasks, regression tasks, both}

3. [1] List three phonemes whose production involves nasal resonances.

4. [1] Name or describe one phoneme that exists in some human language but not in English.

5. [1] What is argmin $_{w}$ (editDistance(w, *alphet)*), where w $\in$ {*outfit, alphabet, artistic}*?

6. [1] Which of the following language models would be able to detect the error in: *This is a great class to talk* (where the intended word was *take*)?
   a) Unigram model
   b) Bigram model
   c) Trigram model
   d) None of the above

/21

7. [4] Match

Regression models ___	a. output numeric values
Classification models ___	b. output novel creations
Generative models ___	c. output a category or a probability distribution over categories
Pretrained models ___	d. output features useful for various tasks

6. [3] How would each of the following pairs of components be ordered in a typical spoken dialog system? Fill in each blank with one of {before, after, simultaneously, either before or after}

a) natural language understanding _____ lexical disambiguation

b) speech synthesis _____ dialog management

c) speech recognition _____ user intent classification

7. [2] Which two of the following are sequence-to-sequence problems?

a) text classification
b) machine translation
c) information retrieval
d) speech synthesis

8. [1] Fill in the blank. "A context-free grammar consists of a set of rules or productions, each of which expresses the ways that the symbols of the language can be _____ and a lexicon of words and symbols."

a) Grouped
b) Ordered
c) Grouped and ordered
d) None of the above

9. [2] Briefly explain one reason to compute the cosine in natural language processing.

10. [2] What is the difference between lemmatization and stemming?

11. [3] Vector-space representations and logic-based representations both have their advantages. Of the following tasks, name one where a vector-space representation would be clearly better, and one where a logic-based representation would be clearly better. {tweet sentiment analysis, machine translation, chatbot response generation, information extraction, information retrieval, spam filtering, detecting toxic comments, inferring music genres from the lyrics}.

12. [4] Any given phoneme, say /i/ or /t/, can have variant pronunciations.
    a) Give two reasons why phoneme pronunciations can vary.
    b) Explain how speech recognizers can (usually) cope with such variation.

13. [2] Recalling Chris Mendoza's talk, what is "model distillation" and how can it be useful in search engines?

14. [3] When trying to extract information from a LLM, if zero-shot prompting doesn't give the answer you are hoping for, how can you get a better result? Explain and illustrate.

15. [5] Very early speech synthesizers worked by simply concatenating sounds obtained by lookup. For example, the nonce phrase *global warmification* could be synthesized by concatenating the sounds for *global*, *warm*, *ific*, *a*, and *tion.* Name two ways in which this is inadequate, and for one of them, explain how modern synthesizers do better.

16. [3] Language models, traditionally, were just systems able to estimate the likelihood of any given word sequence. Explain why this basic ability is enough to support most of the fancy things that modern "large language models" can do.

17. [2] Why is endpointing harder than just detecting silence?

18. [3] In the sentence *the poor are always with us*, what part of speech is the word *poor*? Give two reasons for your answer.

19. [4] Explain why the following statement is partly true and partly false: “For training a large language model, it doesn’t matter if some small fraction of the data is bad (incorrect, ungrammatical, hateful, confused, etc.).”

20. [6] tf-idf weighting was invented for information retrieval, but is useful more generally. Pick one of the following and explain why and how both tf and idf are useful: {machine translation, sentiment detection, summarization}

21. [4] Imagine you are maintaining a speech recognizer, and a potential customer, a car dealer, tells you “in our trial, your system worked well, except the word *Ford* was frequently misrecognized as *fod.* If you can fix this, we’ll choose your product.” To fix this would you change the language model or the acoustic model? Why?

22. [3] What is Wizard-of-Oz testing? When would you use it?

23. [10] A euphemism is a politely worded phrase covering for a harsher meaning, like *Karl is no longer with* us for *Karl's dead.* It can be humorous to see a frank phrase juxtaposed with a creative euphemism. Design a system to create such juxtapositions in the domain of teaching assessment. Given a corpus of anonymous student comments, like

> *taught dated topics, was often unprepared and disorganized, belittled the weaker students, unclear assignments were frustrating, dressed like a mad scientist, really cared for the students, grading was ridiculously lenient*

and a corpus of teacher self-evaluation phrases, like

> *used active learning techniques, adopted a spontaneous lecturing style, made the class engaging for even the weaker students, focused on the foundational concepts, structured the assignments to enable creative solutions, fostered student success*

your system should be able to output gems of sarcastic humor like

> "When a professor says he *structured the assignments to enable creative solutions,* the students probably would say just *the assignments were unclear."*

> "When a professor thinks *the students were captivated by my quirky sense of humor*, the students are probably saying *none of his jokes were funny."*

Explain how you would build this, describing the role and design of each software component and the training and evaluation methods. Describe any additional data you would seek to use, but be aware that there exists no corpus of euphemisms used in sarcastic humor.

24. [6] As Manning-Raghavan-Schuetze note, "in a search engine whose scoring was based on term frequencies, a web page with numerous repetitions of "*maui golf real estate*" would rank highly for the query "maui golf real estate". Modern search engines avoid this problem. Pick two of the below and briefly explain its contribution, or explain why it doesn't help.
    a. It's because we usually take the *log* when computing `tf` .
    b. It's because taking the cosine effectively normalizes for document length.
    c. It's due to something else that modern search engines do.