

# **Final Exam**

## **601.467/667 Introduction to Human Language Technology**

Fall 2020

Johns Hopkins University

Co-ordinator: Philipp Koehn

15 December 2020

Complete all questions.

Use additional paper if needed.

Time: 180 minutes.

Name of student: \_\_\_\_\_

## Q1. Dialog Systems

20 points

1. Which (can be none or more than one) of the following are easy for today's conversational agents: [2 pts]

- (a) question answering
- (b) decision support
- (c) meeting booking
- (d) chitchat

*As easy is a relative term, the only one that's clearly not easy is decision support, so (c), (a,c), or (a,c,d) are all correct*

2. Write down three ethical concerns / considerations regarding conversational agents (a.k.a. chatbots): [5 pts]

*1. bias in data; 2. gender of agents; 3. privacy; 4. controlled generation (e.g. not racist / sexist); 5. access for all (e.g. more than just US, China, Japan) ; 6. equitable design; 7. safety (e.g. vehicles / mental health)*

3. Describe 3 differences between task and non-task driven chatbot evaluation [3 pts]

*1. tasks we can see if task is fulfilled 2. we can measure slot filled 3. many semantic choices in non-task 4. generally no automatic metrics for chat at the moment 5. chat is more concerned about engagement/rapport and not minimizing the length of the conversation*

4. True or False, if False explain why [10 pts]

- (a) Machine translation is important for Chatbot Evaluation

*False - most chatbot systems are monolingual. Some of the evaluation methods from MT are useful for chatbot evaluation, but that's about it*

- (b) Information Retrieval is important for chatbots.

*True*

- (c) Current trends show that chatbots are in decline.

*False, chatbots are growing in usage as the technology gets better*

- (d) Chitchat is useless and shouldn't be studied.

*False, chitchat is useful for engagement/entertainment/ rapportbuilding negotiation*

- (e) Companies are all using neural methods now.

*False - companies are largely using template and rule based methods because of controllability*

- (f) Entity linking is important for persona.

*False, entity linking is more important for contextual information*

- (g) Mutual information is important for neural generative chatbot models since this method takes into account the probability of the source given the target and the probability of the target given the source.

*True*

- (h) The mutual information objective decreases the likelihood of so-called dull responses.

*True*

- (i) Chatbots are attempting to include world knowledge, personal knowledge, and persona. This can only be done with transformer models.

*False, seq2seq models can do this, but most SOA are indeed using transformers*

- (j) Rule-based chatbots are a good choice when starting a project.

*True*

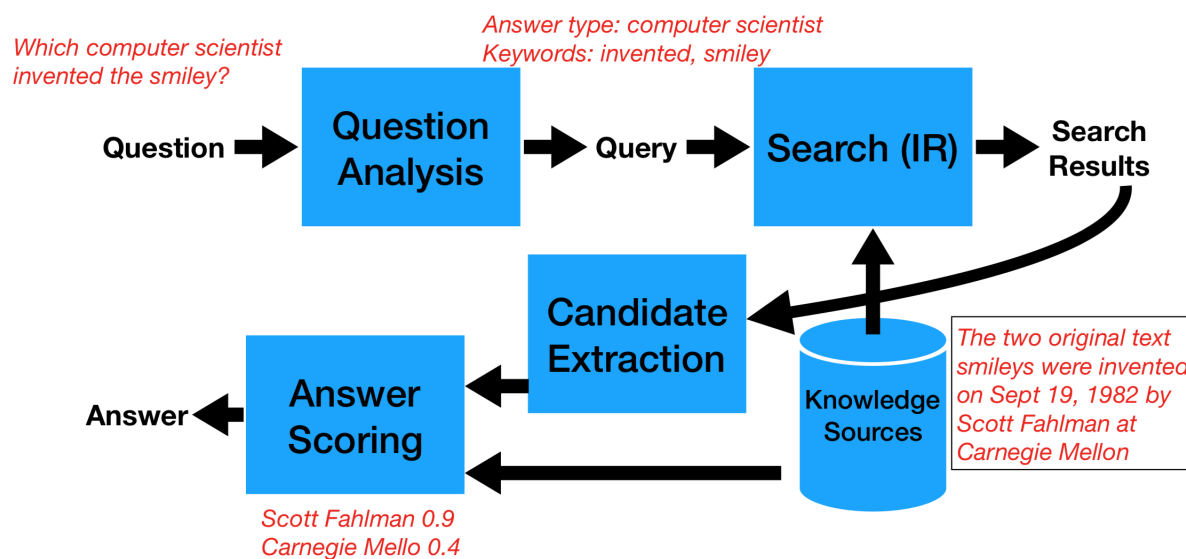
## Q2. Question Answering

20 points

1. Question: Below is a canonical system diagram for factoid question answering. Describe in a sentence each what is the purpose of each component:

*This is a relatively open-ended freebie question. Any answer related to the following would work.*

- (a) Question Analysis: *Figure out what should be the type for the answer (e.g. person or date) and what are the relevant keywords*
- (b) Search (IR): *Find candidate documents from a large knowledge source*
- (c) Candidate Extraction: *Extract potential answers from each document*
- (d) Answer Scoring: *Rank the answers by aggregating evidence over multiple sources and using more advanced models.*



2. Explain the similarities and differences in the problem definition of Question Answering, Machine Reading Comprehension, and Information Retrieval. (3-4 sentences are sufficient.)[10pt]

*This is another freebie question. Any reasonable response that demonstrates some understanding of the different fields is fine. Below is an example.*

*They are all about answering some question/query in some form.*

*But IR focuses on finding documents (one of which may contain the answer), while QA and MRC tends to return shorter answer segments. QA also differs from IR in that the answer is often a summarization based on multiple documents.*

*The knowledge base can differ: IR and QA tend to have large document collections, while MRC focuses on answering based on a single document.*

*This reflects the difference in motivation: IR/QA are application-focused while MRC focuses on modeling reasoning.*

### Q3. Digital Humanities

20 points

1. How does collaboration with humanities scholars relate to more familiar research in science and industry? [8 pts]

*The common presence of "knowledge workers", individuals responsible for understanding, explaining, and acting within specific and esoteric domains. Humanities lack the support structure (funding, research, etc) of science and industry, which exacerbates the need for machine learning techniques that can be employed and interpreted without dedicated computational expertise.*

2. What are the trade-offs between human reasoning and computational techniques? [8 pts]

*Human reasoning, while exceptional at careful scrutiny and creative inference, is deeply influenced by past experience and expectations, and limited in attention and consistency. Computational techniques are relatively limited in their ability to follow intricate patterns of reasoning, connect diverse sources of information, or incorporate "common sense" knowledge, but have limitless patience and attention, and particular biases are easier to avoid at the algorithmic level.*

3. Which format(s) of data is one most likely to encounter working with humanists? [4 pts]

- (a) Plain text files
- (b) SQL databases
- (c) Spreadsheets
- (d) RDF

*Spreadsheets.*

## Q4. Clinical NLP

20 points

1. What is secondary use of medical records? [8 pts]

*Secondary use of clinical data is using it for a different purpose than the one for which it was originally collected. Examples include making medical advances, discovering negative drug interactions, identifying patient cohorts, ensuring patient safety protocols.*

2. Assume you have a dataset of clinical notes (text). You remove all metadata that contains personally identifiable information. Can you publicly release the resulting dataset? Why not? [8 pts]

*The clinical text will still contain personally identifiable information that is protected by HIPAA, such as names, identification numbers, and other personal information about individuals.*

3. Named entity recognition (NER) is an NLP task common across many domains of data, but the types of entities often differ. Which of these entities would a clinical NER system identify? [4 pts]

- ☐ Medications
- ☐ Organizations
- ☐ Locations
- ☐ Citation
- ☐ Treatments
- ☐ Geopolitical entities
- ☐ Symptoms
- ☐ Author names

*Medications, Treatments, Symptoms*

## Q5. Ethics

20 points

### 1. Institutional Review Boards and Informed Consent:

- (a) Describe what Institutional Review Boards are. [5 pts]

*IRB are committees composed of agents from different areas of society (scientists, doctors, patients, etc.) whose purpose is to protect the rights and welfare of human subjects of research conducted in an institution (university, hospital, etc.). This committee evaluates the ethical aspects of human subject research.*

- (b) What is the purpose of the Informed Consent in human subject research? [5 pts]

*The Informed Consent's primary purpose is to ensure that the participants in human subject research are correctly informed about their involvement, rights, and risks in research. The participants must express that they voluntarily consent to the experiment. The researchers must observe specific protocols/means to ensure that the participants understand what is being done in the experiments and understand the document (informed consent) that they're signing. They also have to give participants the right to withdraw consent at any time.*

### 2. Ethical Principles

- (a) Why do we need to ensure that AI algorithms take into account the principle of justice? How could considering fairness in AI affect the algorithm's output? [5 pts]

*We need to ensure that AI algorithms consider the principle of justice to promote prosperity, preserve solidarity, and avoid unfairness. Otherwise, the algorithms could be biased and take unfairly discriminatory decisions based on demographic (gender, race, age, for instance) or geographic (zip code) aspects of the individuals that could affect their well-being (salary, job promotion, parole, etc.). To ensure that the algorithm is more fair, engineers might need to remove some input information (demographic or geographic, for instance) that could derive in lower accuracy during the validation of the model but also less discriminatory decisions.*

- (b) Explain the concept of 'Explicability' in AI ethical principles and how it is related to the other four principles.<sup>1</sup> [5 pts]

*The AI systems must be transparent and reproducible (making the code available, if possible). Ideally, we should be able to explain what all the parts of the algorithm are calculating and how the algorithms take decisions. However, as that is not always possible (a neural network can be quite opaque), we must analyze and understand the algorithms' decisions/outputs at every moment and assess if their actions are beneficial, fair, maleficent, etc. Auditing the algorithms is a possibility. In this manner, we will be able to evaluate if the other four principles (beneficence, autonomy, non-maleficence, and justice) are observed or not.*

---

<sup>1</sup>This question is referred to the five principles we saw in class, as explained in: Floridi L, Cowls J (2019) A unified framework of five principles for AI in society. Harvard Data Science Review.