Dialog Systems Exam Sample Questions
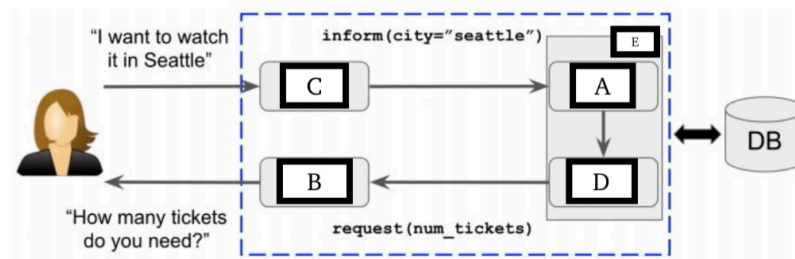
July 10, 2025

Figure 1: Task-oriented Dialog System

1. Examine the task-oriented dialog system in Figure 1. Assume the system uses text and not speech.

   (a) Write the letter of the corresponding component next to each name:

   - (POL) Dialog Policy: __D__
   - (NLU) Natural Language Understanding: __C__
   - (DST) Dialog State Tracking: __A__
   - (NLG) Natural Language Generation: __B__
   - (DM) Dialog Manager: __E__

   (b) Write the initials (e.g., DM for Dialog Manager) of the dialog component next to the statement that best matches it:

   - Keeps track of the users goals: ____
   - Historically involved sentence planning and surface realization: ____
   - Proposes an assignment of values to slots from an ontology: ____
   - Typically learned through reinforcement learning and decides what action to take next: ____

2. Name and provide examples of five types of dialog system applications.

   1. **Task-oriented**:
      - Example: Booking a flight, ordering food, scheduling appointments.
   2. **Open-domain conversation**:
      - Example: Casual conversation, customer support, mental health chatbots.
   3. **Information-seeking**:
      - Example: Answering questions about weather, news, or general knowledge.
   4. **Personal assistant**:
      - Example: Setting reminders, managing calendars, controlling smart home devices.
   5. **Educational**:
      - Example: Tutoring systems, language learning assistants, interactive quizzes.

3. What is turn-taking in dialog and why is it difficult for a dialog system to handle it? Name and explain at least two reasons.

4. How is speech different from text for dialog systems? Name five differences.

> Natural speech is very different from written text:
> 1. **ungrammatical**
> 2. **restarts, hesitations, corrections**
> 3. **overlaps**
> 4. **pitch, stress**
> 5. **accents, dialect**

5. Name and explain five types of speech acts.

> 1. **Assertive**: Speaker commits to the truth of a proposition.
>   - Example: "It's raining outside."
> 2. **Directive**: Speaker wants the listener to do something.
>   - Example: "Stop it!"
> 3. **Commissive**: Speaker commits to doing something themselves.
>   - Example: "I'll come by later."
> 4. **Expressive**: Speaker expresses their psychological state.
>   - Example: "Thank you!"
> 5. **Declarative**: Speaker performs actions using performative verbs.
>   - Example: "You're fired!"

6. Given the following exchange between A and B, which of Grice's maxims is most clearly broken? What is the implicature of this broken maxim?

   (a) Person A: Did you study enough for the exam?

   (b) Person B: I was so busy yesterday. I spent almost the entire day cleaning my house!

Circle the maxim and describe the implicature below:

<div align="center">Quantity, Quality, Relation, Manner</div>

7. Describe **grounding** in dialog systems and name at least two grounding signals, each briefly explained in a sentence.

>

8. What are anaphora and cataphora, and why are they problematic for dialog systems?

9. What is adaptation/entrainment in dialog, and in what forms does it take place?

10. You have developed a function to measure lexical entrainment over the turns in a conversation. You apply this to a dataset of human–machine conversations with your system and find that over time, the entrainment increases. What does this tell you about your system?

11. What is the difference between a chatbot and a dialog system?

12. What is an ontology, and how are the ontology slots distinguished from one another?

13. Explain and give examples for the two semantic concepts in dialog acts.

> 1. **intent or dialogue act type**:
>   - encodes the system or the user intention in a (part of) dialogue turn
> 2. **semantic slots and values**:
>   - further describe entities from the ontology that a dialogue turn refers to

14. Name and explain three challenges in NLU.

15. Name and explain two methods for using ASR hypotheses for the NLU inputs. Explain why one is more robust than the other. Explain the ways for NLU prediction using ASR hypotheses.

16. Give an example of a word confusion network for ASR and explain its benefit.

17. Explain how NLU can be formulated as a sequence-to-sequence learning task.

18. What is MLM (Masked Language Modeling)?

19. Name two challenges for modular dialog system architecture and explain the hybrid approach as its solution.

20. What is delexicalization, and why is it useful?

21. What are some challenges associated with dialog state tracking? Name at least four.

22. Explain the two main approaches to dialog state tracking.

23. What are the automatic and human metrics that can be used for NLG? (Name three for each and explain.)

24. What is the interpretation of perplexity as an NLG metric?

25. What is the main difference between BERTScore and other metrics, such as ROUGE and METEOR, in NLG evaluation?

26. Explain three things one must consider when evaluating a dialog system with volunteers.

27. Name and explain three evaluation factors to be included in the questionnaire of human evaluators for task-oriented and chat-oriented systems.

28. What are the advantages and disadvantages of crowd-sourcing? Name two for each.

29. How does a data-driven automatic dialog measure work, and what are its problems?

30. What are the benefits and drawbacks of using prompting for dialog evaluation?

31. What are the two main features of an ideal evaluation metric?

32. What is the antibiotic effect in the context of dialog evaluation metrics?

33. Name and explain the factors that make an NLG system good.

34. What is the pros and cons of a template-based natural language generator?

35. Name and explain the components of the trainable generator pipeline (by Walker).

36. How does class-based language modeling work for NLG?

Classes:
- *inform_area*
- *inform_address*
- *inform_phone*
- *request_area*
- ...

Generation process:
1. Generate utterances by sampling words from a particular class language model in which the dialogue act belongs to.
2. Re-rank utterances according to scores.

37. How does an RNN network improve NLG over its predecessors? In other words, what are its main features?

- An RNN (Recurrent Neural Network) as a language generator improves over its predecessors due to its natural ability to model sequences.

- Main features:
  ‣ **Handling Long-Term Dependencies**
  ‣ **Flexibility to Condition on Auxiliary Inputs**

38. How does LSTM prevent the vanishing gradient problem?

- Consider memory cell, where recurrence actually happens:

$$C_t = i_t \odot \hat{C}_t + f_t \odot C_{t-1}$$

- We can back-propagate the gradient by chain rule:

$$\frac{\partial E_t}{\partial C_{t-1}} = \frac{\partial E_t}{\partial C_t} \frac{\partial C_t}{\partial C_{t-1}} = \left(\frac{\partial E_t}{\partial C_t}\right) f_t$$

39. Why can a semantically conditioned LSTM be better than a non-conditional LSTM?

- A semantically conditioned LSTM can be better than a non-conditional LSTM because it incorporates additional semantic information into the generation process, allowing it to produce more contextually relevant and coherent responses.
- This conditioning helps the model to focus on specific aspects of the dialog state, leading to improved performance in generating appropriate utterances.

40. What distinguishes GPT-3 from earlier GPT models?

- Effectively the same model structure with substantially more parameters and training data:

	GPT	GPT-2	GPT-3
Parameters	117M	1.5B	175B
Data	12GB	40GB	570GB

41. What is a major drawback of GPT, as well as semantically conditioned GPT?

- **Cons**: produces hallucinations

42. What are the three key elements in each dialog turn in dialog management?

43. What are the three main challenges in modeling dialog, and how does defining it as a control problem provide a solution?

1. How to define the state space?
2. How to tractably maintain the dialogue state?
3. Which actions to take?

- **Solution**: Define dialogue as a control problem where the behaviour can be automatically learned.

44. How is dialog management framed in a Markov Decision Process (MDP)?

1. **Data**:
   - Dialogue states
   - Reward: a measure of dialogue quality
2. **Model**:
   - Markov decision process
3. **Predictions**:
   - Optimal system actions

- s_t: dialogue states
- a_t: system actions
- r_t: rewards
- $p(s_{t+1} \mid s_t, a_t)$: transition probability

45. What is the main difference between generative and discriminative models in belief tracking?

- **Discriminative models**: focus on the relationship between observations and states without modeling how the observations were generated.
- **Generative models**: model both the states and the observations, considering how states generate observations. The probability of the state depends on how likely it is that this state generated the given observation, combined with our prior belief about the state itself.

46. Why is exact belief tracking intractable in partially observable MDP-based dialog systems?

- Requires summation over all possible states at every dialogue turn - **intractable!**

47. Name and explain three requirements for belief tracking.

1. **Dialogue history**: The system needs to keep track of what happened so far in the dialogue. This is normally done via the **Markov property**.
2. **Task-orientated dialogue**: The system needs to know what the user wants. This is modelled via the **user goal**.
3. **Robustness to errors**: The system needs to know what the user says. This is modelled via the **user act**.

48. What is the Hidden Information State (HIS) model in dialog systems, and how does it address the challenges of belief tracking in partially observable environments?

- The Hidden Information State (HIS) model is a practical framework for POMDP-based spoken dialogue management.

- It addresses the challenges of belief tracking in partially observable environments by structuring the belief state around several components:
  1. **Observation**: N-best list of user acts
  2. **User Goal**: Partitions of the goal space built according to ontology
  3. **Dialogue history**: Grounding states
  4. **Hypotheses**: Every combination of user act, partition and history

- **Belief state**: Distribution over most likely hypotheses

49. What is the Bayesian Update of Dialogue State (BUDS) model, and how does it improve upon previous approaches like the Hidden Information State (HIS) model in belief tracking?

- The Bayesian Update of Dialogue State (BUDS) model is an advancement in POMDP frameworks for spoken dialogue systems.

- Improve upon previous approaches by:
  ‣ Further decomposes the dialogue state
  ‣ Produces tractable belief state update
  ‣ Transition and observation probability distributions can be parametrized and their shape learned

50. Why is grounding important in open-domain dialog systems?

- Prevent hallucinations and improve factual accuracy
- Improve trustworthiness and transparency
- Respond accurately on dynamic or rare topics

51. What is Retrieval-Augmented Generation (RAG), and how does it work? Name and explain its two main components.

52. What is the key advantage of separating the retriever from the generator in RAG systems?

53. Name two advantages of RAG systems over traditional end-to-end language models.

54. What are the four pipeline stages in a RAG system?

55. What are three limitations of RAG systems?

56. How can retriever quality be improved in RAG systems?

57. What are two techniques to improve alignment between the generator and retrieved knowledge?

58. What are the three key evaluation metrics for grounded dialog?

59. What is the potential risk of RAG systems in open-domain dialog?

60. Name and explain three reasons why open-domain dialog is more challenging than task-oriented dialog.

61. Why are generic responses like 'I don't know' problematic in chatbot conversations, and how can they be mitigated?

62. What are three ways to mitigate hallucination in dialog systems?

63. What is a major risk when a chatbot loses context in multi-turn conversations, and what are two ways to mitigate it?

64. What types of errors fall under pragmatic errors in chatbot responses, and what are three ways to mitigate such errors?

65. How can chatbots mitigate toxic or biased responses?

66. Why is it difficult for chatbots to handle multi-intent utterances like 'Book a flight to Tokyo and what's the weather there?'

67. What distinguishes a retrieval-based language model from a standard language model?

68. Give five reasons why we need retrieval-based language models.

69. What are the three main design questions for retrieval-based language models?

70. What is the main idea behind kNN-LM?

71. Give an advantage and a disadvantage of fine-tuning in adapting an LM to downstream tasks.

72. Give an advantage and a disadvantage of reinforcement learning in adapting an LM to downstream tasks.

73. Give an advantage and a disadvantage of retrieval-based prompting.

74. Name and explain three methods for adapting a retrieval-based LM for downstream tasks.

75. Name and explain four key effectiveness points in downstream tasks for retrieval-based LMs.

76. Name and explain five scenarios when retrieval-based LMs should be used.

77. What is an agent?

78. Name and explain the two competing views on agents.

79. What is the fundamental difference between current and classic agents?

80. Name three types of AI agents and compare them in terms of expressiveness, reasoning and adaptivity.

81. Name and explain three methods that can be used to teach LLMs how to properly use tools.

82. What is the ReAct agent, and what are its benefits?

83. What are the four components of the unified framework for LLM agents?

84. What is multi-agent orchestration, and why do we need it?

85. Give two examples of a multi-agent system and explain how they work.

86. What are two potential risks of multi-agent systems, and why are they difficult to deal with?

87. What is LLM alignment, and why is it important?

88. How does Proximal Policy Optimization (PPO) work, and what are its two benefits?

89. What are three drawbacks of PPO?

90. How does Direction Preference Optimization (DPO) improve upon PPO?

91. What is Reinforcement Learning from Human Feedback (RLHF)?

92. What are the challenges of the *human feedback* in RLHF?

93. What are the challenges of the *reward model* in RLHF?

94. What are the challenges of the *policy* in RLHF?

95. Name and explain the three key concepts for language agents.

96. Why are reasoning and acting helpful for agents?

97. How do LLM agents have short-term and long-term memories, and what are they most useful for?

98. Name and explain three planning paradigms for language agents. Give an advantage and a drawback for each.

99. What are the advantages of code agents?

100. What is the purpose of the Model Context Protocol (MCP)?