1. Which statement is true about the Unigram algorithm for tokenization? 1 point

- It segments text into manageable parts and assigns unique IDs.
- It involves splitting text into individual characters.
- It begins with a large list of possibilities and gradually narrows down based on how frequently they appear in the text.
- It evaluates the benefits and drawbacks of splitting and merging two symbols to ensure its decisions are valuable.

2. Identify the advantages of using data loaders in natural language processing (NLP). Select all that apply. 1 point

- [ ] Splits text into characters to ensure vocabulary is small
- [x] Seamlessly integrates with the PyTorch training pipeline
- [x] Enables shuffling of data
- [x] Enables batching of data

ndividual characters.
ossibilities and gradually narrows down based on how frequently they appear in the text.
drawbacks of splitting and merging two symbols to ensure its decisions are valuable.

lata loaders in natural language processing (NLP). Select all that apply. 1 point
ensure vocabulary is small
ıe PyTorch training pipeline