1. **Training loss / accuracy curves for vanilla RNN and LSTM training**
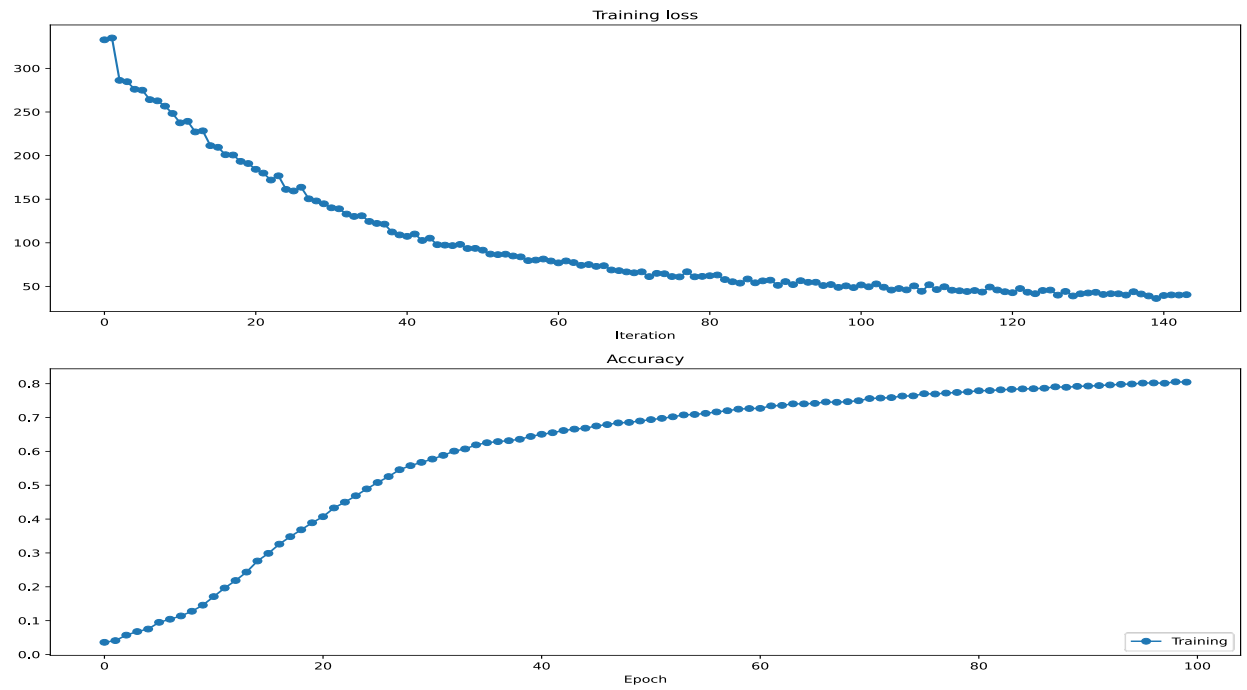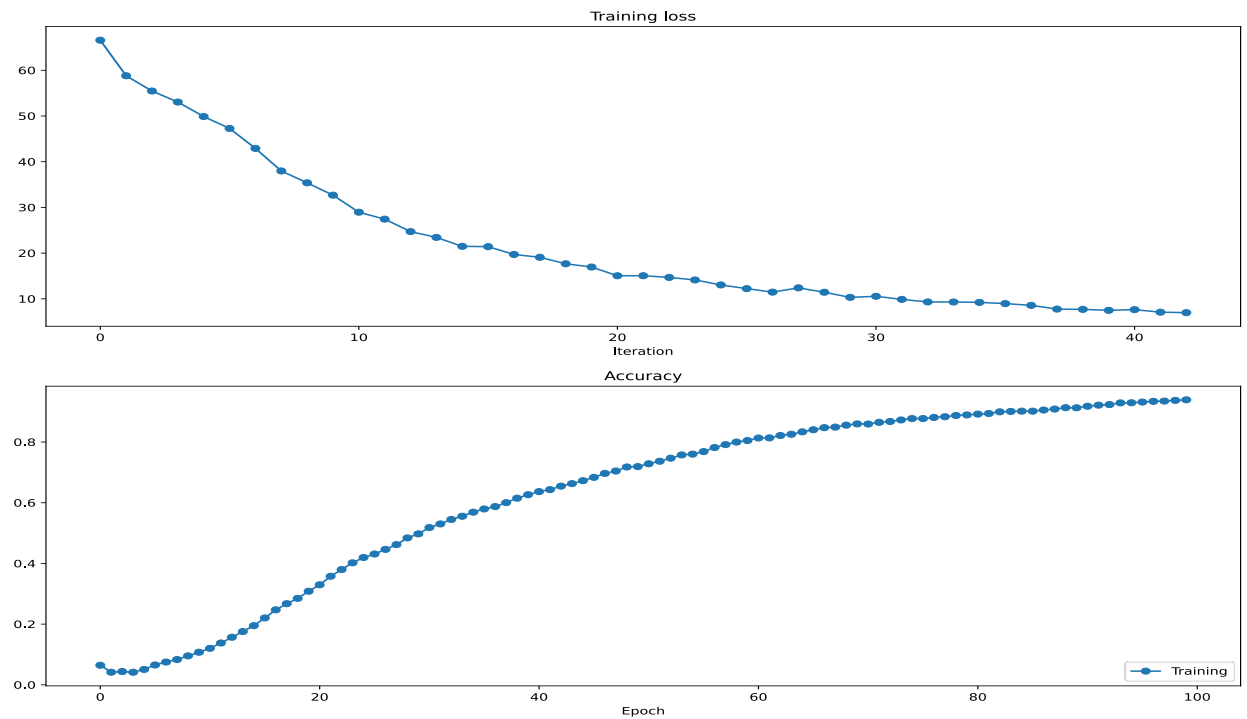
LSTM:

RNN:

2. **Sample text generation from a trained model**

LSTM:
she was too slippery; and when she was coming to, but it was nothing else to be seen: she had tired herself out with trying, the rabbit actually took to herself, 'after such a little glass box that was nothing

RNN:
she had never forgotten that, if you drink much from a bottle marked 'poison,' it is almost certain to get out of that dark hall, and wander about among those beds of bright flowers and those cool fountains

3. **Answers to inline questions about recurrent net behavior**

While playing around with different parameters, I observed two patterns. First one which is bit obvious is that the LSTM performs better with larger timesteps but is not so good with smaller timesteps. However, for Vanilla RNN, the behavior is reverse i.e. Vanilla RNN performs worse with the larger timesteps but does a good job with smaller ones. I think this is primarily because of the vanishing gradient issue which happens with deeper networks in Vanilla RNN.

The second thing I observed in both the models (but more in LSTM) is repetition of text when I increase the test length. I suspect this is happening because of lack of randomization. We may be able mitigate this issue by including randomization through dropout layers. Also, instead of using top choice as our predicted word which results in extremely repetitive and predictable text, we can treat the output of the model as a probability distribution and sample from that to simulate more to the real-world behavior.