Chapter 12: Translation as Encoding and Decoding

1. TRUE/FALSE - Machine translation (MT) evolved in the manner which is typical of AI systems over the past 50 years, from symbolic rules to statistical models to neural models.
2. TRUE/FALSE - The basic architecture for Google Translate (and similar systems) consists of the coupling of one recurrent neural network for encoding a sentence with respect to the source language with another recurrent neural network, informed by the "source" network, for decoding the sentence with respect to the target language.
3. TRUE/FALSE - For longer sentences, the basic recurrent neural network architecture tends to "forget" earlier parts of the sentence, resulting in rather poor performance. In the late 1990s, a research group in Switzerland proposed a solution to this problem: the individual units in a recurrent neural network should have a more complicated structure, with specialized weights that determine what information gets sent on at the next time step and what information can be "forgotten." These researchers called the more complex units "long short-term memory" (LSTM) units.
4. TRUE/FALSE - Automated machine translation in the deep-learning age is a triumph of big data and fast computation. To create a pair of encoder-decoder networks to translate from, say, English to French, the networks are trained on more than thirty million human-translated pairs of sentences. Deep recurrent neural networks made up of LSTM units trained on large data collections have become the bread and butter of modern natural-language processing systems, not just the encoding and decoding networks used by Google Translate, but also for speech recognition, sentiment classification, and question answering.
5. What does MM think of the claim that machine translation is now close to "human level"? (Please provide at least some of the evidence that she provides in justifying her thinking on the issue.)

Mitchell believes that the claim that machine translation is now close to "human level" is unsupported by proper evidence. Firstly, the statistics ranking the success of machine translation rely on averages, which are not a good statistical measure due to the effect of outliers on the average calculation. Next, Mitchell argues that the studies only look at translation of sentences rather than longer passages, which are more complex and rely more on semantics and contextualizing. Lastly, the studies use passages that lack ambiguity/complexity – typically documents from news stories or Wikipedia – which ignores the problems machines face when dealing with ambiguity (204-205).

6. What does a caption generating system do?

A caption generating system outputs a phrase describing an input image (208).

7. TRUE/FALSE - The architecture of an image captioning system differs from that of a machine translation system in that the encoding RNN of the latter is replaced by a CNN in the former.
8. TRUE/FALSE - Automated image captioning suffers from the same kind of bipolar performance seen in language translation. When it's good, it seems almost magical. But its errors can range from slightly off to completely nonsensical.
9. TRUE/FALSE - While Microsoft's CaptionBot says it can "understand the content of any photograph," the problem is that the opposite is true. Even when their captions are correct, these systems don't understand photos in the sense that humans understand them. In fact, they tend to be incapable of describing the most interesting aspects of a photo, the way it speaks to us, to our experience, emotions, and knowledge about the world. That is, it misses the *meaning* of the photo.