1. Neural networks in general (10% for BS and MS students)

a) Artificial neuron

The artificial neuron is the smallest unit of neural networks. What it does is that it takes multiple inputs and multiplies them by weights, adds a bias, and sends these calculations through an activation function to decide the output. Neural networks are made of many connected artificial neurons, learning patterns from data ("Transformer Models," slides "A Few Concepts (with Short Explanations)," & "The Artificial Neuron").

b) Loss function (or cost function)

A loss function will measure how wrong a neural networks predictions are. It compares the predicted output to the correct answer and gives a number that tells us the eror. The loss is sent back through the network and the output is used to change the weights.

So if the loss is high, the network is making bad predictions. If the loss is low, the network is preforming well. The goal of this is to minimize the loss functions so the network improves. ("Transformer Models," slide "A Few Concepts (with Short Explanations)").

c) Backpropagation

Backpropagation is how a neural network learns from its mistakes. It makes a prediction first and checks how wrong it is using the loss function. Then it sends this error backward through the network to update the weights. This helps the network improve by using the gradient descent method, until it makes better predictions ("Transformer Models," slides "Backpropagation" & "What's Backpropagation Doing?").

d) Hidden layer

A hidden layer is a layer in a neural network between the input and output layers. It helps the network learn by processing the input data before passing it forward. The input layer takes in raw data and the hidden layer finds patterns or features in that data.

Every hidden layer contains nodes that perform specific tasks, like recognizing edges in an image or detecting certain colors. These nodes pass their findings to the next layer, enabling the network to learn intricate data representations.

Hidden layers are what make deep learning models powerful by allowing them to break down and process data step by step ("Transformer Models," slide "A Few Concepts (with Short Explanations)"; Coursera).

4. Transformer models and HuggingFace (30% for BS students – 25% for MS students)

First model:
XLMr-ENIS-QA-Is is an Icelandic reading comprehension Q&A model. So it can find answers to questions in Icelandic and English. It is built on XLM-Roberta which is a multilingual model that understands many languages.

This model follows an encoder-only structure, what that means is it processes text by looking at both directions at once to understand the meaning. How it was trained was that it translated English datasets and they were used along with the Natural Questions in Icelandic dataset.

Second model mbert-finetuned-ner:

The “ner” stands for Named Entity Recognition which is what this model is used for. IT means it can find words like names of people, places, companies, and other important entities in text. It is built on mBERT which is a version of BERT that works in many languages.

The model follows an encoder-only structure. It was trained on datasets for NER, which allows it to recognize and classify important words in different languages. This makes it useful for tasks like finding names in news articles or recognizing locations in documents.

Third mode M2M100 418MI:

This model is used for machine translation, which means it can translate text between different languages. It is built on Facebook's M2M100 model, which can handle translation between over 100 languages without using English as a middle step.

The model follows an encoder-decoder structure, where the encoder first processes the input text and the decoder generates the translation. It was trained on a huge multilingual dataset collected from the internet, which makes it capable of translating directly from one language to another. For example like Icelandic to French or Spanish to Chinese.