

Answer Key

| Question Number | Question | Answer |
|-----------------|---|-----------------|
| 1 | What layer does Attention use to learn weight probabilities | Softmax |
| 2 | Attention can be computed on single strings | True |
| 3 | This layer type can apply a model multiple times to an input | TimeDistributed |
| 4 | Keras models can include whole Keras models as layers | True |
| 5 | Networks of networks can be combined to make predictions and are called | Ensembles |
| 6 | Ensemble networks have smaller numbers of parameters than single networks | False |