[Lecture] LSTM is dead. Long Live Transformers! Seattle Deep Learning YT video, Nov 2019  
<https://www.youtube.com/watch?v=S27pHKBEp30>

[Article, technical] Hyperparameter optimization for Neural Networks, by NeuPy, 2016 (haven’t read yet)  
<http://neupy.com/2016/12/17/hyperparameter_optimization_for_neural_networks.html>

[YT Channel] Siraj Raval, lots of good stuff on machine/deep learning.  
<https://www.youtube.com/c/SirajRaval/playlists>

Automatic Machine Learning  
<https://www.youtube.com/watch?v=jn-22XyKsgo>

[StackExchange] Example image and info on Bayesian optimisation (haven’t read yet)  
<https://datascience.stackexchange.com/questions/8043/optimizing-parameters-for-a-closed-black-box-system>  
  
[Lecture] Lecture 16.3 — Bayesian optimization of hyper-parameters, Colin Reckons.  
<https://www.youtube.com/watch?v=con_ONbhD2I>

[Paper] Pervasive Attention: 2D Convolutional Neural Networks for Sequence-to-Sequence Prediction <https://arxiv.org/abs/1808.03867>

[YT video] References on Code emporium YT video: Transformer Neural Networks - EXPLAINED!

REFERENCES  
[1] The main Paper: <https://arxiv.org/abs/1706.03762>  
[2] Tensor2Tensor has some code with a tutorial: https://www.tensorflow.org/tutorials/...  
[3] Transformer very intuitively explained - Amazing: http://jalammar.github.io/illustrated...  
[4] Medium Blog on intuitive explanation: https://medium.com/inside-machine-lea...  
[5] Pretrained word embeddings: https://nlp.stanford.edu/projects/glove/  
[6] Intuitive explanation of Layer normalization: https://mlexplained.com/2018/11/30/an...  
[7] Paper that gives even better results than transformers (Pervasive Attention): <https://arxiv.org/abs/1808.03867>  
[8] BERT uses transformers to pretrain neural nets for common NLP tasks. : https://ai.googleblog.com/2018/11/ope...  
[9] Stanford Lecture on RNN: http://cs231n.stanford.edu/slides/201...  
[10] Colah’s Blog: <https://colah.github.io/posts/2015-08>...

[Github] simple Bayesian optimiser toolkit from fmfn  
<https://github.com/fmfn/BayesianOptimization>

[Github] Hyperopt package for Tree structured Parzen estimator

[article] A Conceptual Explanation of Bayesian Hyperparameter Optimization for Machine Learning  
<https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f>

[Paper] DARTS: Differentiable Architecture Search (2018)  
<https://arxiv.org/abs/1806.09055>

[Lecture YT vid] EvoDevo: Deep Neural Network Hyperparameter Optimization wtih Genetic Algorithms <https://www.youtube.com/watch?v=giBAxWeuysM>

[article] Automated Machine Learning Hyperparameter Tuning in Python*: A complete walk through using Bayesian optimization for automated hyperparameter tuning in Python (TPE)*  
<https://towardsdatascience.com/automated-machine-learning-hyperparameter-tuning-in-python-dfda59b72f8a>

<https://machinelearningmastery.com/what-is-bayesian-optimization/>

<https://github.com/thuijskens/bayesian-optimization>  
<https://youtu.be/jtRPxRnOXnk>

[Youtube video] Bayesian Hyperparameter Optimization for Keras (8.4) [ for neural networks, check this out]  
<https://www.youtube.com/watch?v=sXdxyUCCm8s>

[YT video, lecture] Machine Intelligence - Lecture 18 (Evolutionary Algorithms), Kimia Lab  
<https://www.youtube.com/watch?v=3-NiZPbkr7A>