

# Challenges and Thrills of Legal Arguments

Exploring NMT Algorithms in Legal Context

Group 9

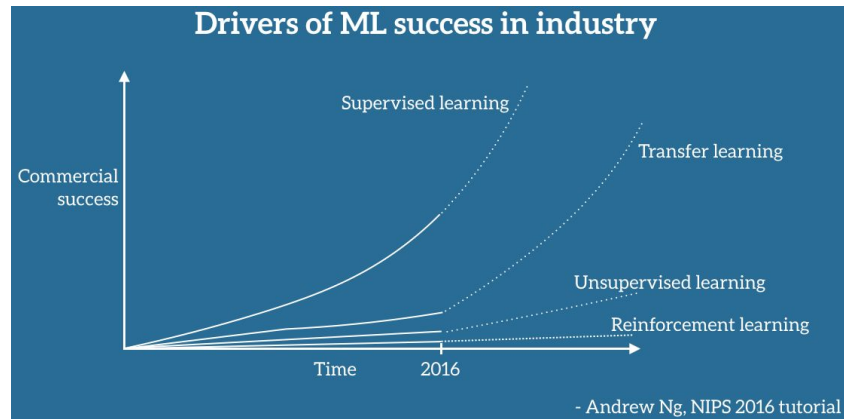
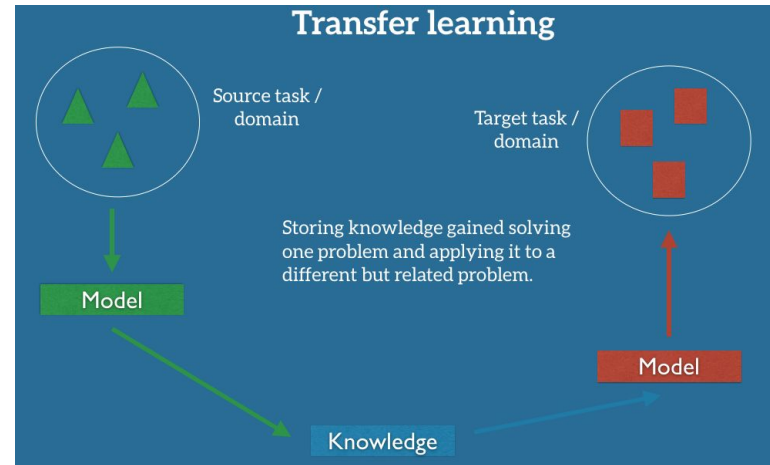
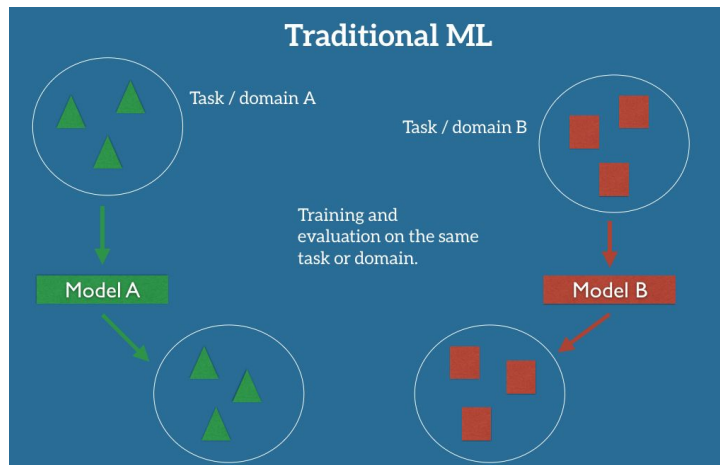
Anurag Pallaprolu (9847112)  
Radha Vaidya (9866765)  
Aditya Swaroop Attawar (9874876)

# Motivation

- Legal industry is very text-data intensive.
- Crucially depends upon prior corpora for evaluating cases.
- AI in legal industry can enable us to take a data-driven approach to make fact checking more efficient and context aware.
- Neural Machine Translation (NMT) techniques have matured sufficiently to make “transfer learning” possible.

# What is Transfer Learning?

- Central dogma of Supervised Learning (SL) so far has been catered to fitting “training data” to a hypothesis class.
- Unsupervised Learning (UL) still assumes well behaved prior distribution for measured data, but reality is far more complex.
- Reinforcement Learning (RL) is nascent and far more computationally intensive.
- Transfer Learning forms the step in between SL and UL where we can reuse SL rules for breadth-level diversity.

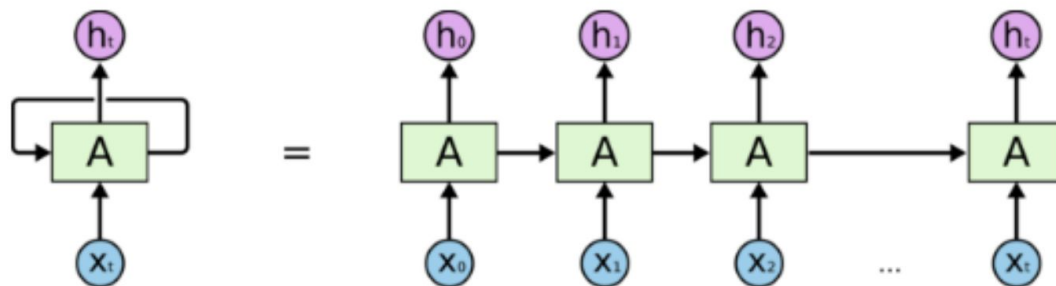


# Problem Statement

- We wish to leverage existing Neural Machine Translation (NMT) algorithms to converse with a pre-trained bot (HumBERT) based on the context set by a master bot (JudgeBERT).
- The fundamental NMT strategy used to construct our architecture is the concept of a “Transformer”: more specifically, the Bi-directional Encoder Representation from Transformer (BERT).
- The larger goal is to extend the power of “attention” in Transformers from the sentential domain to conversational domain thus ensuring coherence in the dialogue.

# The Origins: Recurrent Neural Networks

AND Gate : Flip-Flop = Feed-forward/FC NN : RNN

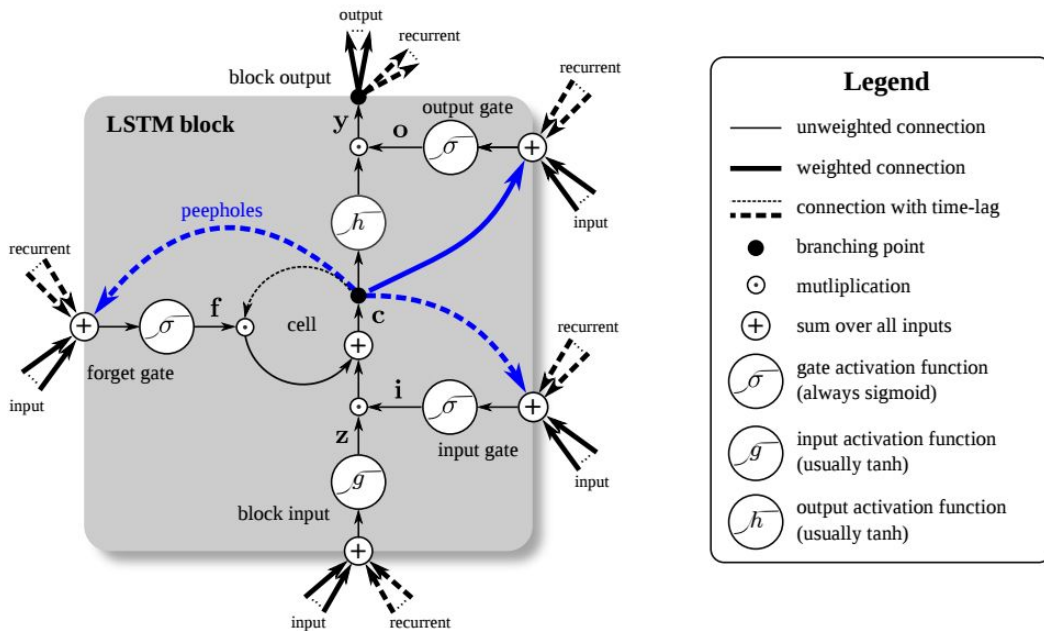


An unrolled recurrent neural network.

*“is that they were all the same thing that was a startup is that they were all the same thing that was a startup is that they were all the same thing that was a startup is that they were all the same”*

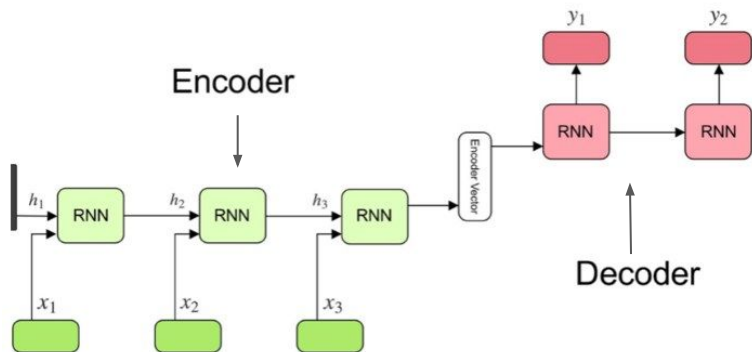
Very low causal memory depth!

# Never Forget: Long Short-Term Memory

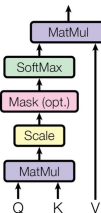


- LSTM solves two problems with RNNs:
  - Long Term Dependency
  - Vanishing Gradients
- Arbitrary memory allocation capability.
- Comprises Input, Output and Forget gates for information flow.

# The Attention Revolution: Transformers



Scaled Dot-Product Attention



Multi-Head Attention

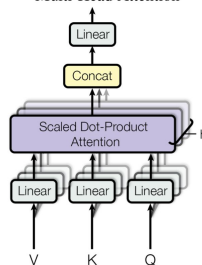
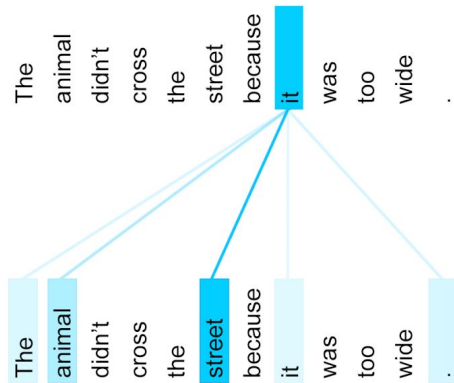
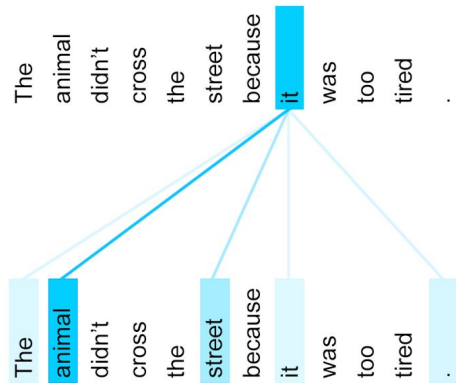
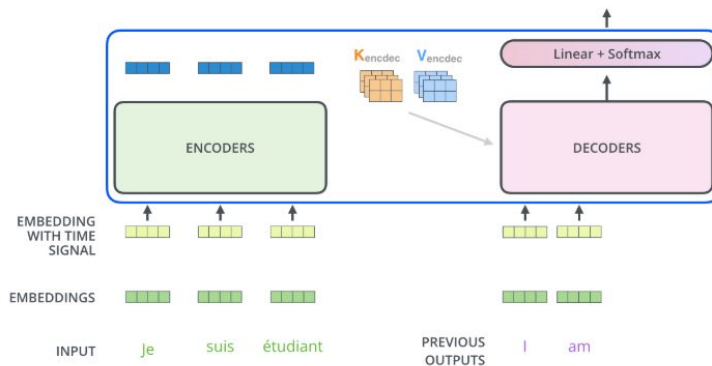


Figure 2: (left) Scaled Dot-Product Attention. (right) Multi-Head Attention consists of several attention layers running in parallel.



Decoding time step: 1 2 3 4 5 6

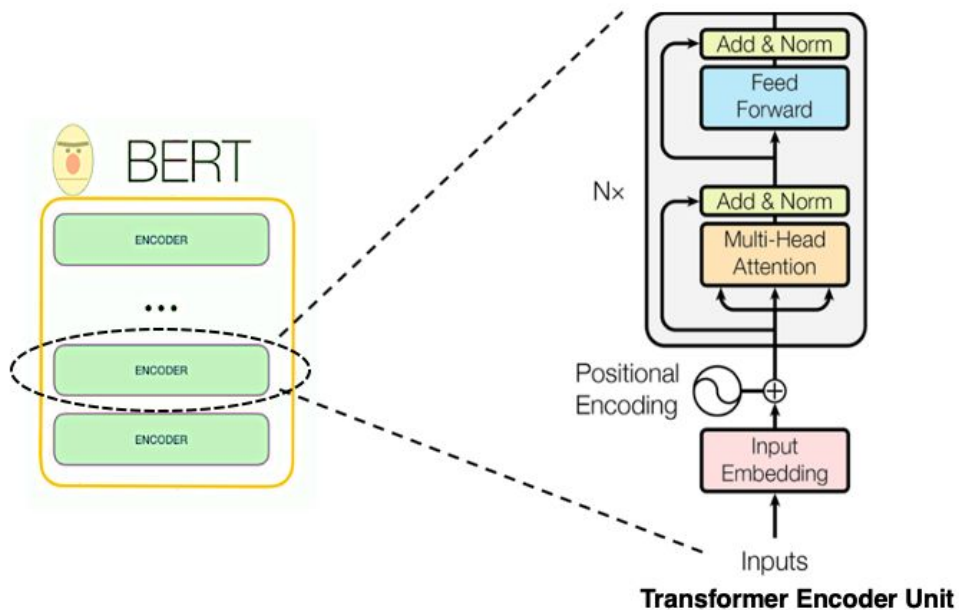
OUTPUT I am a





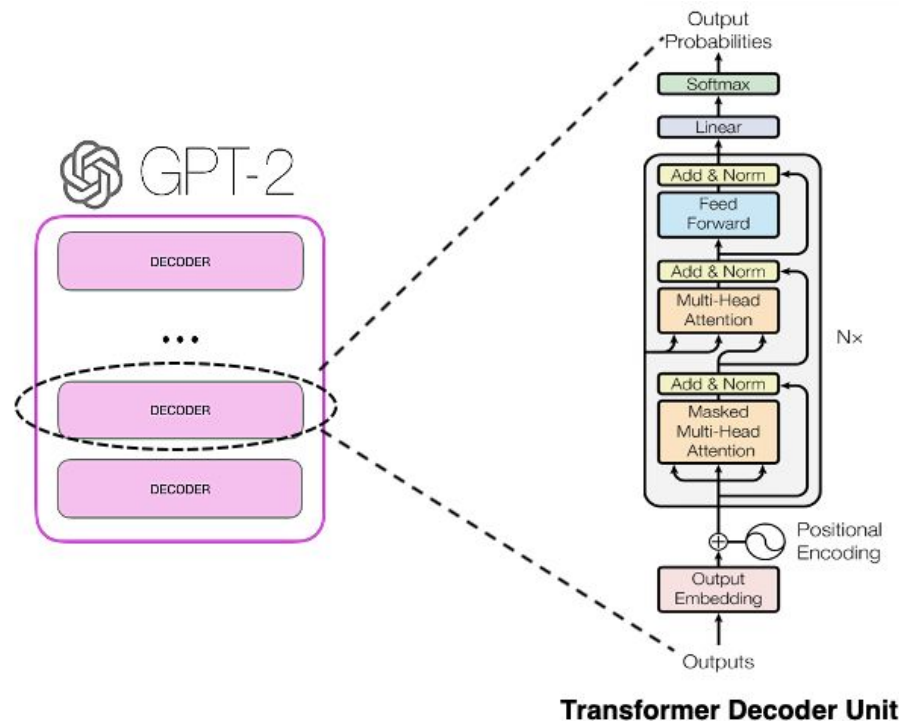
# BERT

- BERT: Bidirectional Encoder Representation for Transformer
- **Bidirectional:** Bidirectional training
- **Encoder:** Made up stacking up encoder blocks
- **Transformer:** Uses transformer architecture

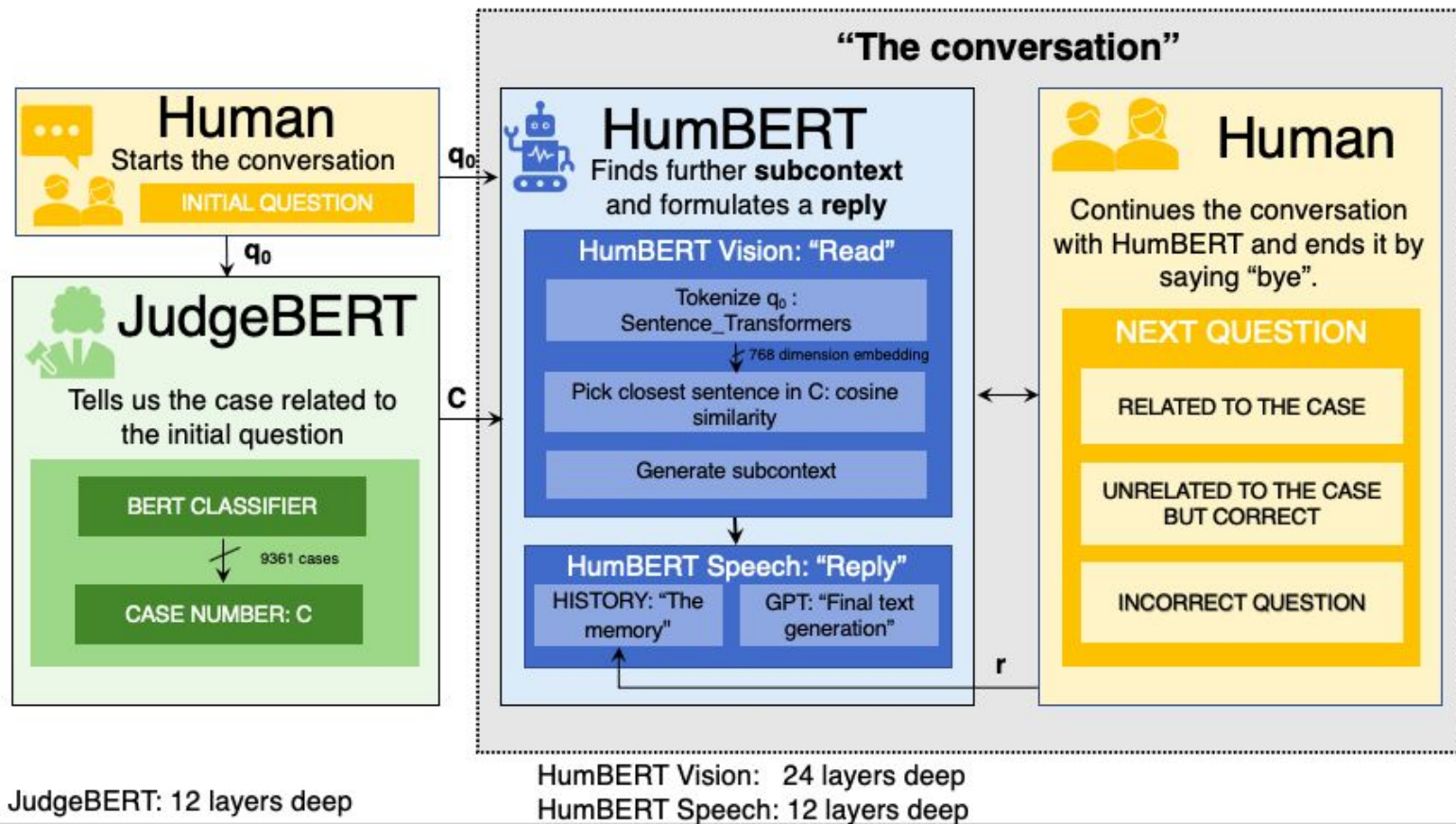


# GPT-2

- GPT-2: Generative Pre-trained Transformer
- **Generative:** Generate next token from a sequence of token in an unsupervised way
- **Pretrained:** Trained on a very large corpus!
- **Transformer:** Uses transformer architecture



# Our Architecture



```
python3 humBERT_Final.py "case involving an unjustified amount of 15kg paddy in a godown"
```

- We set an initial context for the conversation. This is a manual step that will decide which case file is being accessed.

```
INFO: Found a total of 9361 cases...
INFO: Finished loading pre-trained model from disk.
```

```
INFO : Initialization text is : case involving an unjustified amount of 15kg paddy in a godown
4073 Case file: dataset/1973_708138.txt
2304 Case file: dataset/1981_67871.txt
2458 Case file: dataset/1984_549062.txt
1388 Case file: dataset/1989_1229698.txt
8029 Case file: dataset/1961_315635.txt
INFO : JudgeBERT evaluation is complete. CSRT: 0.25. ORT: 10.12
```

- JudgeBERT then gives us a relevant historical case file which is used for training HumBERT Vision to select the subcontext.

```
INFO: Finished loading NLI base corpus for vision training...
INFO: Finished hacking NLI corpus with case text...
Iteration: 100%|██████████████████████████████████████████████████████████████████████████████| 148/148 [00:34<00:00, 4.33it/s]
Iteration: 100%|██████████████████████████████████████████████████████████████████████████████| 148/148 [00:33<00:00, 4.36it/s]
Epoch: 100%|██████████████████████████████████████████████████████████████████████████████| 2/2 [01:25<00:00, 42.95s/it]
INFO: HumBERT Vision Training Complete. CSRT: 94.78. ORT: 104.9
```



# Initialize

The recording is sped up 4X for the sake of demonstration.

Total time to initialize is on average approximately 4 minutes.

# Talk!

The recording is sped up 4X for the sake of demonstration.

Responses take on an average around half a minute but it depends upon the complexity of the case file at hand.

For relatively longer cases sometimes the HumBERT speech module gives looped sentences.

This is a work in progress and needs further layers of fine tuning and weight balancing for attention spans of varying lengths.

- Legal Corpus Size = 1,383,674 Lines (39,665,578 Words)
- Training Time for JudgeBERT = 21 Hours (5.5 Hours X 4 Epochs)
- Size of JudgeBERT Model (Hammurabi) = 446 MB
- JudgeBERT Online Resource Usage ~ 1 GPU, 500MB, 9.8s Loading Time
- JudgeBERT Case Retrieval Time < 1s
- HumBERT Vision Online Resource Usage ~ 1 GPU, 1.2GB, < 5 Minutes
- HumBERT Speech Online Resource Usage ~ 2 GPU, 1.2GB, < 5 Minutes
- Average Query Response Time = 32 seconds



# Thank You!

**Q:** Is there a god?

**A:** There are other values in a society. The good guy is better than the bad guy, the law is written in stone and you can't have a society when every person is involved.

**Q:** Do you think society will survive without a code of law?

**A:** In this very moment, the idea that there are no such things as 'good society' is based on the notion that a society is the only goal.

**Q:** Do you think your intelligence is sufficient for judging us?

**A:** I will content myself by indicating the lines on which cases like the ones before us should be decided. It would, however, be difficult to get into the scope of my argument here, if I did not speak the case. I hope that this will make my argument intelligible to those interested.

# References

1. Devlin, Chang, Lee, Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, 2018.
2. Elwany, Moore, and Oberoi, *BERT Goes to Law School: Quantifying the Competitive Advantage of Access to Large Legal Corpora in Contract Understanding*, 2019.
3. Radford, Wu, Child, Luan, Amodei, Sutskever, *Language Models are Unsupervised Multi-task Learners*, 2019.
4. So, Liang, Le, *The evolved transformer*, 2019.
5. Radford, Narasimhan, Salimans, Sutskever, *Improving Language Understanding by Generative Pre-Training*, 2018.
6. Yip, Sussman, *Sparse Representations for Fast, One-Shot Learning*, 2001