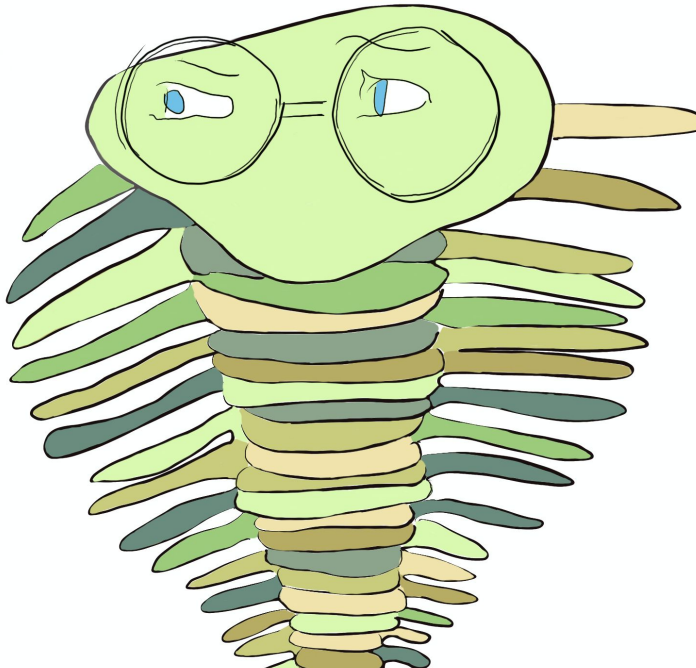
Deep Learning Model Architectures for **Natural Language Processing**

Powerful, Efficient Processing with Artificial Neural Networks

Jon Krohn, Ph.D.



jonkrohn.com/talks

github.com/jonkrohn/DLTFpT

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. Recurrent and Advanced Neural Networks

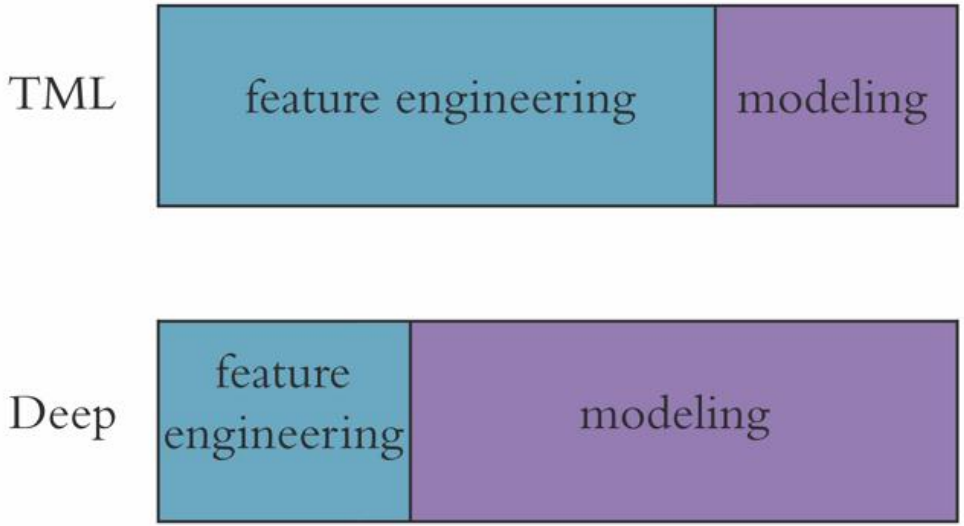
Deep Learning for NLP

1. **The Power and Elegance of Deep Learning for NLP**
2. Modeling Natural Language Data
3. Recurrent and Advanced Neural Networks

Two Core Concepts

1. Deep Learning
2. Natural Language Processing (NLP)

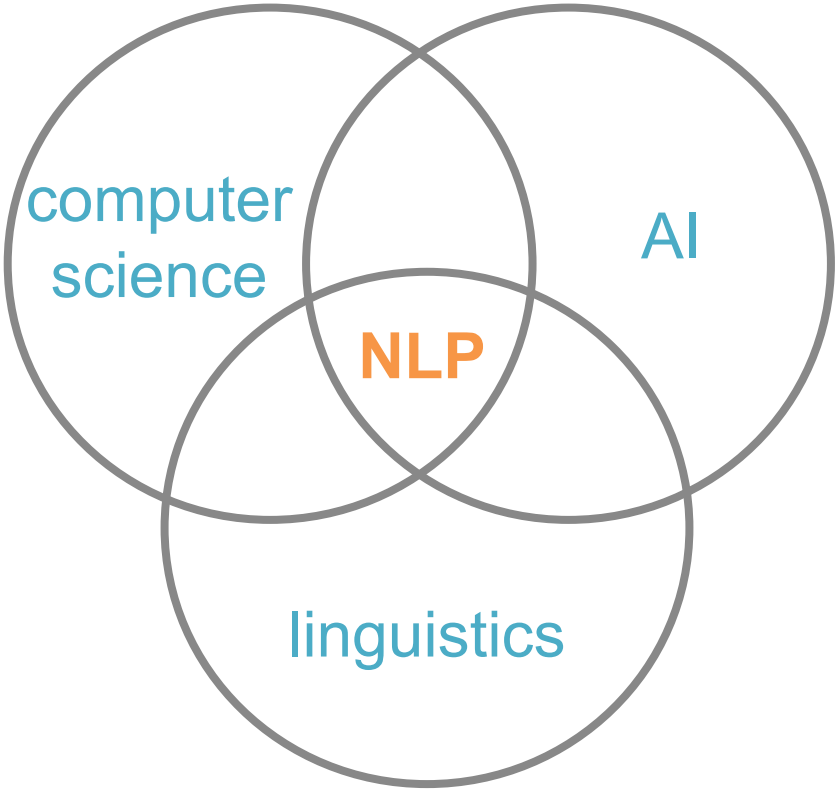
Traditional ML vs Deep Learning



Two Core Concepts

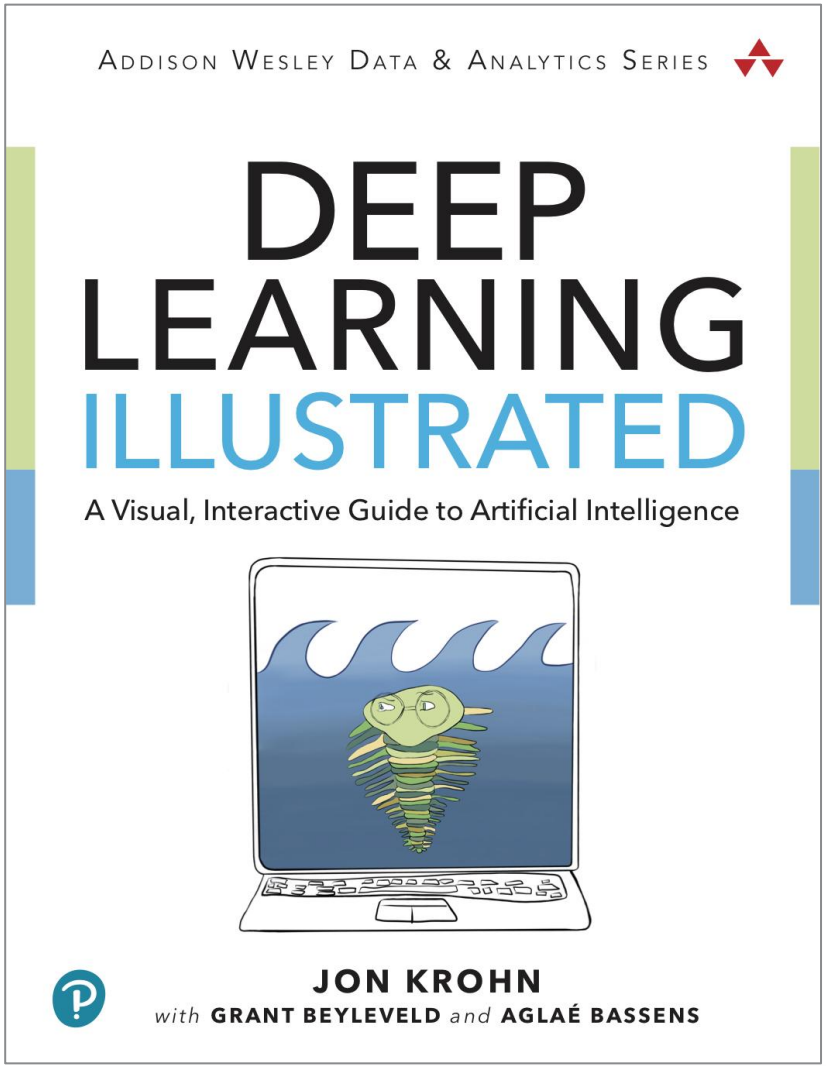
1. Deep Learning
2. Natural Language Processing (NLP)

Natural Language Processing



Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. **Modeling Natural Language Data**
   - Document Classification with a Dense Net
   - Ditto with a ConvNet
3. Recurrent and Advanced Neural Networks



Deep Learning

Theory Assumed:

- **Artificial neurons**
- **Cost functions**
- **Stochastic gradient descent**
- **Dense and convolutional layers**
- **Pooling**
- **Dropout**

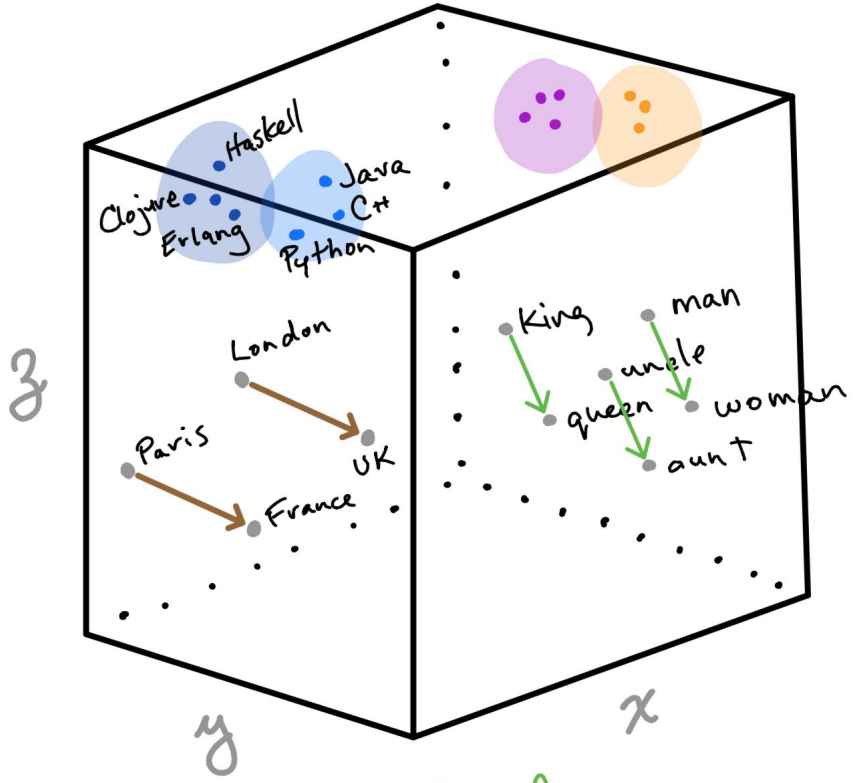
35% off orders:

bit.ly/iTkrohn

(use code KROHN

during checkout)

Vector Representations of Words



n-dimensional space

Word Vector Arithmetic

$$v_{\text{king}} - v_{\text{man}} + v_{\text{woman}} = v_{\text{queen}}$$

$$v_{\text{bezos}} - v_{\text{amazon}} + v_{\text{tesla}} = v_{\text{musk}}$$

$$v_{\text{windows}} - v_{\text{microsoft}} + v_{\text{google}} = v_{\text{android}}$$

$$v_{\text{cu}} - v_{\text{copper}} + v_{\text{gold}} = v_{\text{au}}$$

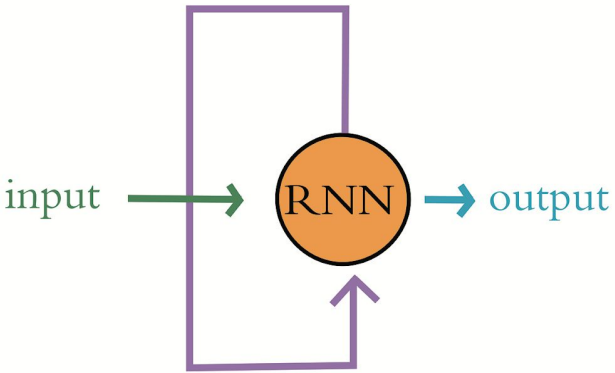
Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
    - Recurrent Neural Networks
    - LSTMs
    - Bi-Directional LSTMs
    - Stacked LSTMs
    - Parallel Network Architectures
    - Transfer Learning in NLP

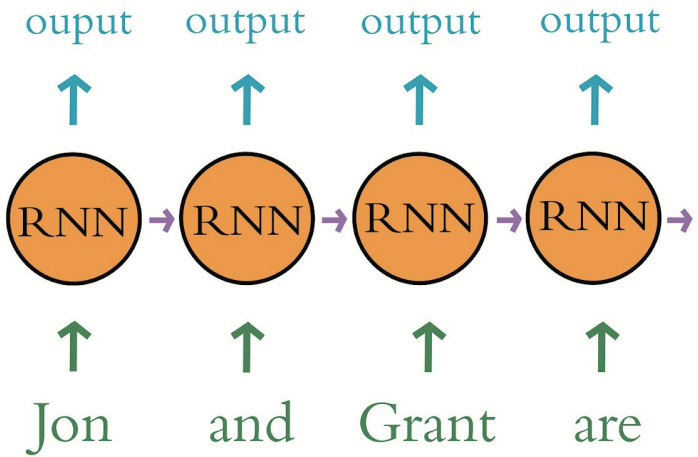
Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
    - Recurrent Neural Networks
    - LSTMs
    - Bi-Directional LSTMs
    - Stacked LSTMs
    - Parallel Network Architectures

Recurrent Neural Network



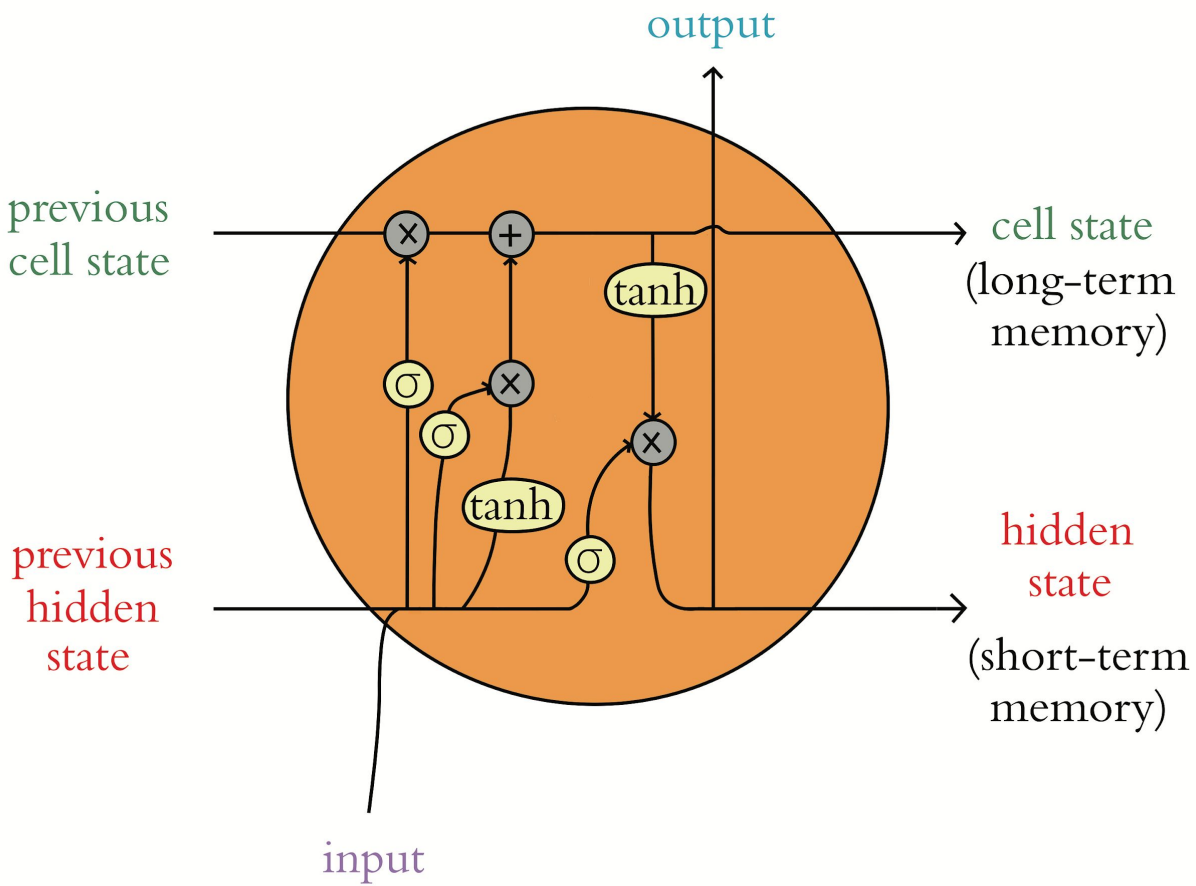
RNN unpacked



Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
    - Recurrent Neural Networks
    - LSTMs
    - Bi-Directional LSTMs
    - Stacked LSTMs
    - Parallel Network Architectures
    - Transfer Learning in NLP

Long Short-Term Memory Unit



Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
    - Recurrent Neural Networks
    - LSTMs
    - Bi-Directional LSTMs
    - Stacked LSTMs
    - Parallel Network Architectures
    - Transfer Learning in NLP

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
    - Recurrent Neural Networks
    - LSTMs
    - Bi-Directional LSTMs
    - Stacked LSTMs
    - Parallel Network Architectures
    - Transfer Learning in NLP

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
    - Recurrent Neural Networks
    - LSTMs
    - Bi-Directional LSTMs
    - Stacked LSTMs
    - Parallel Network Architectures
    - Transfer Learning in NLP

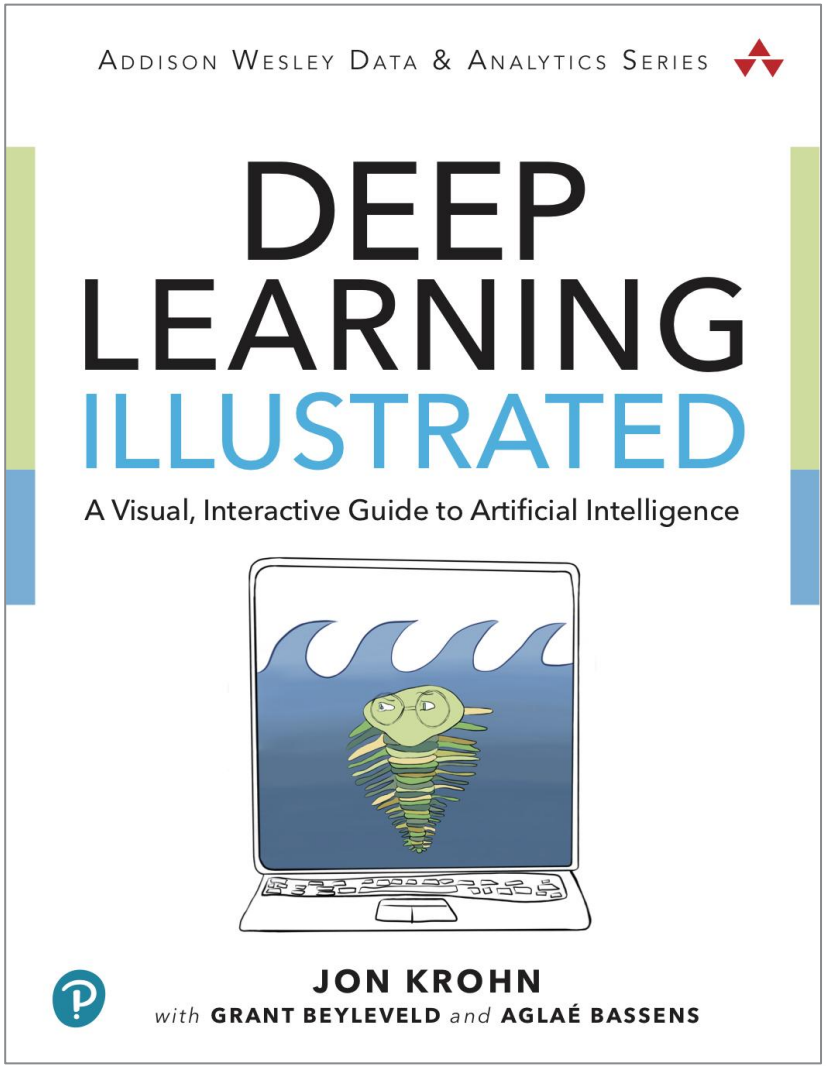
Model	ROC AUC (%)
Dense	92.9
Convolutional	96.1
Simple RNN	84.9
LSTM	92.8
Bi-LSTM	93.5
Stacked Bi-LSTM	94.9
GRU	93.0
Conv-LSTM	94.5
Multi-ConvNet	96.2

Deep Learning for NLP

1. The Power and Elegance of Deep Learning for NLP
2. Modeling Natural Language Data
3. **Recurrent and Advanced Neural Networks**
    - Recurrent Neural Networks
    - LSTMs
    - Bi-Directional LSTMs
    - Stacked LSTMs
    - Parallel Network Architectures
    - Transfer Learning in NLP

Transfer Learning in NLP

- ULMFiT
- ELMo
- BERT
  - RoBERTa
  - DistilBERT
- GPT-2: `talktotransformer.com`
- GPT-3
  - Absolutely massive
  - No transfer learning required for some tasks



35% off orders:

bit.ly/iTkrohn

(use code KROHN during checkout)

Stay in Touch

jonkrohn.com to sign up for email newsletter

linkedin.com/in/jonkrohn

YouTube jonkrohn.com/youtube

twitter.com/JonKrohnLearns