



Home Page

Title Page

Contents



Page 1 of 23

Go Back

Full Screen

Close

Quit

Transformers

Sumantra Dutta Roy

Department of Electrical Engineering
Indian Institute of Technology Delhi
Hauz Khas, New Delhi - 110 016, INDIA.

<http://www.cse.iitd.ac.in/~sumantra>
sumantra@ee.iitd.ac.in

“over-the-hill researcher”[pepsi.mp4]



Home Page

Title Page

Contents



Page 2 of 23

Go Back

Full Screen

Close

Quit

Attention

~~Attention~~ Is All You Need!
Tension

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

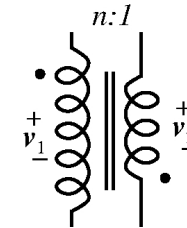
Llion Jones*
Google Research

Aidan N. Gomez*[†]
University of Toronto

Lukasz Kaiser*
Google Brain

https://upload.wikimedia.org/wikipedia/en/1/1c/Transformers_Rise_of_the_Beasts.jpg

– Transformers: Rise of the Beasts



Electrical Transformer

- Translation: sequence translation

[Home Page](#)[Title Page](#)[Contents](#)[◀](#) [▶](#)[◀](#) [▶](#)[Page 3 of 23](#)[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

Next Word Prediction

- There is a structure/order in randomness
- Letters/words do not appear at random
- 1913: **Markov:** $p(\text{letter})$ depends on prior letters
- 1951: **Shannon:** Pred of entropy: English letters
- 1960s: N. Bhattacharyya, Ph.D. student of Mahalanobis. Statistical analysis of Tagore's Gitanjali
- Till 2012: n -grams. Prediction in mobiles
- Prediction based on the last 2 or 3 words $p(y|\mathbf{x})$
- Prediction: words, weather, financial prediction
- 'Knowing' the weather (not climate!), 'knowing' the market (portfolio, high freq trading)
- What is 'knowing' a language well?