

Hierarchical Neural Story Generation

Unveiling the Future of AI Narratives

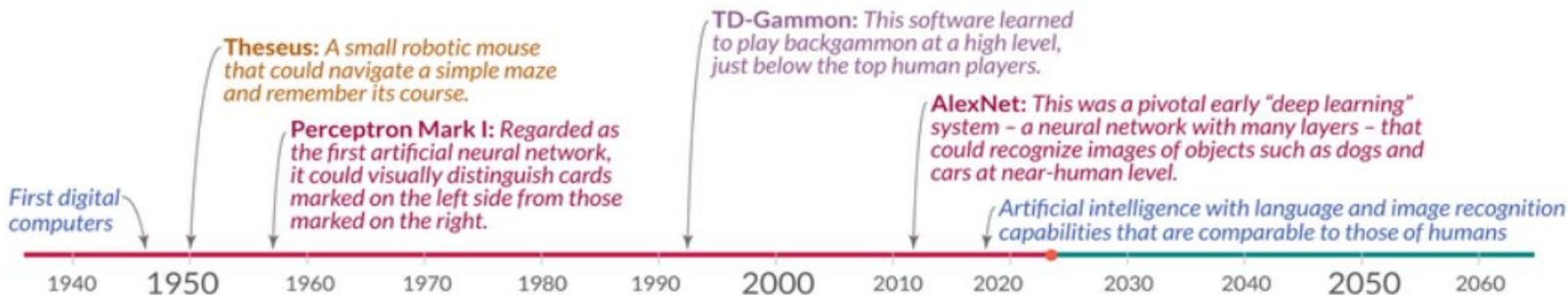


Introduction to Story Generation in AI

- Briefly describe the evolution of story generation in AI.

have achieved state of the art performance on a variety of text generation tasks, such as machine translation and summarization;

A timeline of notable artificial intelligence systems





- As the quality of natural language generated by artificial intelligence systems improves, writing interfaces can support interventions beyond grammar-checking and spell-checking, such as suggesting content to spark new ideas.

Timeline of images generated by artificial intelligence

These people don't exist. All images were generated by artificial intelligence.

Our World
in Data

2014



Goodfellow et al. (2014) - Generative Adversarial Networks

2015



Radford, Metz, and Chintala (2015) - Unsupervised Representation Learning with Deep Convolutional GANs

2016



Liu and Tuzel (2016) - Coupled GANs

2017



Karras et al. (2017) - Progressive Growing of GANs for Improved Quality, Stability, and Variation

2018



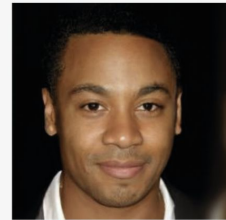
Karras, Laine, and Aila (2018) - A Style-Based Generator Architecture for Generative Adversarial Networks

2019



Karras et al. (2019) - Analyzing and Improving the Image Quality of StyleGAN

2020



Ho, Jain, & Abbeel (2020) - Denoising Diffusion Probabilistic Models

2021 Image generated with the prompt: "a couple of people are sitting on a wood bench"



Ramesh et al. (2021) - Zero-Shot Text-to-Image Generation (OpenAI's DALL-E 1)

2022

Image generated with the prompt: "A Pomeranian is sitting on the King's throne wearing a crown. Two tiger soldiers are standing next to the throne."



Saharia et al. (2022) - Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding (Google's Imagen)

Explaining a Joke

I will explain these jokes:

(1) The problem with kleptomaniacs is that they always take things literally.

Explanation: This joke is wordplay. Someone who "takes things literally" is someone who doesn't fully understand social cues and context, which is a negative trait.

But the definition of kleptomania is someone who literally takes things.

(2) Always borrow money from a pessimist. They'll never expect it back.

Explanation: Most people expect you to pay them back when you borrow money, however a pessimist is someone who always assumes the worst, so if you borrow money from them, they will expect that you won't pay them back anyways.

Input: I was supposed to start writing the paper at 5:00 PM. But then I started playing with this cool new language model for 10 minutes. 10 minutes later, it's suddenly 9:30 PM!

Model Output: This joke is about how time flies when you're having fun. The person was supposed to start writing a paper at 5:00 PM, but they started playing with a language model instead. They thought they were only playing with it for 10 minutes, but it was actually 4.5 hours.

Input: I was going to fly to visit my family on April 6th. My mom said "Oh great, your step dad's poetry reading is that night!" So now I'm flying in on April 7th.

Model Output: The joke is that the speaker's mother is trying to get them to go to their step dad's poetry reading, but the speaker doesn't want to go, so they are changing their flight to the day after the poetry reading.

Input: Reading "War and Peace" doesn't scare me. I once had to review the CL for my intern's whole project on his final day!

Model Output: 'War and Peace' is a very long book, so reading it is a daunting task. However, the speaker is saying they had to review the code for their intern's project on their final day, which is even more daunting.

Input: I think I understand why me and my partner are such a good match. I get excited when I smell food and become super alert when the doorbell rings. My girlfriend Shelley becomes ecstatic when she sees a person who she just saw 5 hours ago.

Model Output: This joke is about dogs. Dogs get excited when they smell food and when the doorbell rings. The girlfriend is excited to see someone she just saw 5 hours ago, which is a trait of dogs.

Input: I tried 10,000 random restarts of my neural network, but I was accused of overfitting. I guess no good seed goes unpunished.

Model Output: This joke is a pun. A neural network is a computer program that can learn from data. A "seed" is a number that is used to initialize a random number generator. A "good seed" is a number that produces a good random number generator. The phrase "no good deed goes unpunished" means that if you do something good, you will be punished for it.

Input: What's the difference between a zebra and an umbrella? One is a striped animal related to horses, another is a device you use to stop the rain from falling on you.

Model Output: This joke is an anti-joke. The joke is that the answer is obvious, and the joke is that you were expecting a funny answer.

The rise of artificial intelligence over the last 8 decades: As training computation has increased, AI systems have become more powerful

The color indicates the domain of the AI system: ● Vision ● Games ● Drawing ● Language ● Other

Shown on the vertical axis is the training computation that was used to train the AI systems.

10 billion petaFLOP

Computation is measured in floating point operations (FLOP). One FLOP is equivalent to one addition, subtraction, multiplication, or division of two decimal numbers.

100 million petaFLOP

The data is shown on a logarithmic scale, so that from each grid-line to the next it shows a 100-fold increase in training computation.

1 million petaFLOP

10,000 petaFLOP

100 petaFLOP

1 petaFLOP = 1 quadrillion FLOP

10 trillion FLOP

100 billion FLOP

1 billion FLOP

10 million FLOP

100,000 FLOP

1,000 FLOP

10 FLOP

The first electronic computers were developed in the 1940s

Pre Deep Learning Era
Training computation grew in line with Moore's law, doubling roughly every 20 months.

Deep Learning Era
Increases in training computation accelerated, doubling roughly every 6 months.

1940 1950

1960 1970 1980 1990
1956: The Dartmouth workshop on AI, often seen as the beginning of the field of AI research

2000 2010 2020
1997: Deep Blue beats world chess champion Garry Kasparov

The data on training computation is taken from Sevilla et al. (2022) – Parameter, Compute, and Data Trends in Machine Learning. It is estimated by the authors and comes with some uncertainty. The authors expect the estimates to be correct within a factor of two.
OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the authors
Charlie Giattino, Edouard Mathieu, and Max Roser



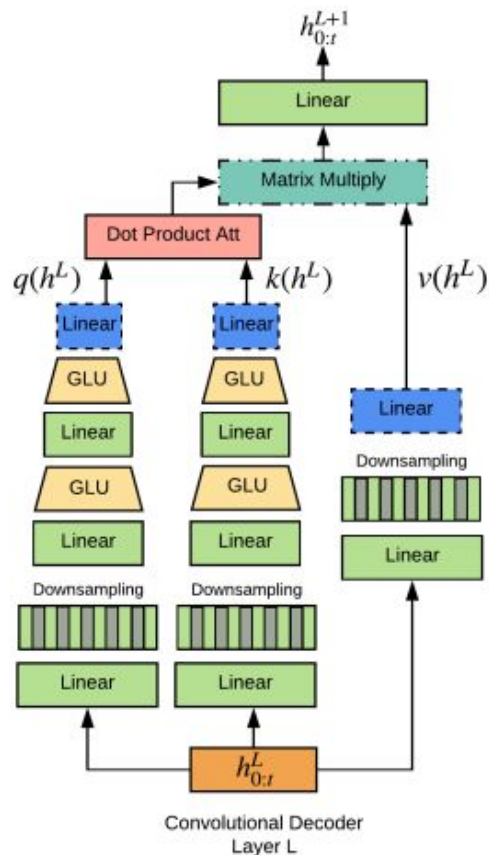
Understanding Hierarchical Models

- Defining hierarchical neural networks: multi-layered networks where each layer processes information at a different level of abstraction.

1. Artificial neural networks are used in the deep learning form of machine learning;
2. It's called deep learning as models use the 'deep', multi-layered architecture of an artificial neural network;
3. As each layer of an artificial neural network can process data, models can build an abstract understanding of the data.
4. This architecture means models can perform increasingly complex tasks, for example understanding natural language or categorising complex file types.

Ex: Recommendation systems, To power virtual assistance, In automatic feature extraction from raw, unlabelled data

Hierarchical Story Generation



High-level structure

1. language models generate on a strictly-word-byword;

****First, we generate the premise or prompt of the story using the convolutional language model (O prompt dá um esboço da estrutura da história.)**

****Second, we use a seq2seq model to generate a story that follows the premise;**

(O condicionamento no prompt faz com que mais fácil para a história permanecer consistente e também ter estrutura em um nível além de frases únicas.)



Hierarchical Text Generation

challenge:

1. “(...)generating long sequences of text into a hierarchical generation task.”;
2. “(...)use an LSTM to hierarchically learn word, then sentence, then paragraph embeddings, then transform the paragraph embeddings into text. “;
3. “generate a discrete latent variable based on the context, then generates text conditioned upon it.”



The Role of Neural Networks in Storytelling

- Detail types of neural networks (e.g., LSTM for handling long-term dependencies, Transformers for managing context).

What are transformer models?

The transformer (represented by the T in ChatGPT, GPT-2, GPT-3, GPT-3.5, etc.) is the key element that makes generative AI so, well, transformational.

Transformer models are a type of neural network architecture designed to process sequential material, such as sentences or time-series data.

The concept of a transformer, an attention-layer-based, sequence-to-sequence (“Seq2Seq”) encoder-decoder architecture, was conceived in a 2017 paper authored by pioneer in deep learning models Ashish Vaswani et al called “Attention Is All You Need”. Since then, in the realms of AI and machine learning, transformer models have emerged as a groundbreaking approach to various language-related tasks.



Compared with traditional recurrent neural networks (RNNs) and convolutional neural networks (CNNs), transformers differ in their ability to capture long-range dependencies and contextual information.

The transformer “requires less training time than previous recurrent neural architectures, such as long short-term memory (LSTM), and its later variation has been prevalently adopted for training large language models on large (language) datasets,” notes Wikipedia.

From machine translation to natural language processing (NLP) to computer vision, plus audio and multi-modal processing, transformers have revolutionized the field with their ability to capture long-range dependencies and efficiently process sequential data.

Modeling Contributions

- Modeling Long Sequences
 - convolutional models
- Writing a coherent, multi-sentence story
 - gated self-attention
 - multi scale self-attention
- Story should relate to prompt

Prompt: The Mage, the Warrior, and the Priest

Story: A light breeze swept the ground, and carried with it still the distant scents of dust and time-worn stone. The Warrior led the way, heaving her mass of armour and muscle over the uneven terrain. She soon crested the last of the low embankments, which still bore the unmistakable fingerprints of haste and fear. She lifted herself up onto the top the rise, and looked out at the scene before her. [...]

Prompt: The Mage, the Warrior, and the Priest

We had been married for almost a year . And we had never been close. And then he came home and I knew . I loved him . I had always known .

model ignores prompt and generates unrelated story



Fusion Mechanism in Story Generation

- Previous work on story generation has explored seq2seq RNN architectures (Roemmele, 2016), but has focused largely on using various content to inspire the stories.
- “We are not aware of previous work that has used hierarchical generation from a textual premise to improve the coherence and structure of stories.”
- Fusion mechanism: Combining different neural networks to maintain relevance to the prompt

Fusion Mechanism in Story Generation

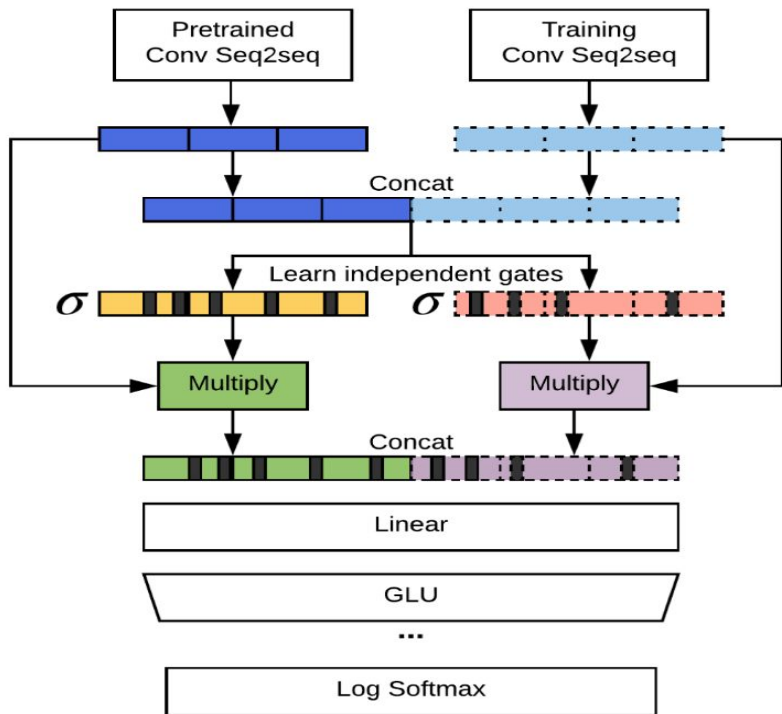
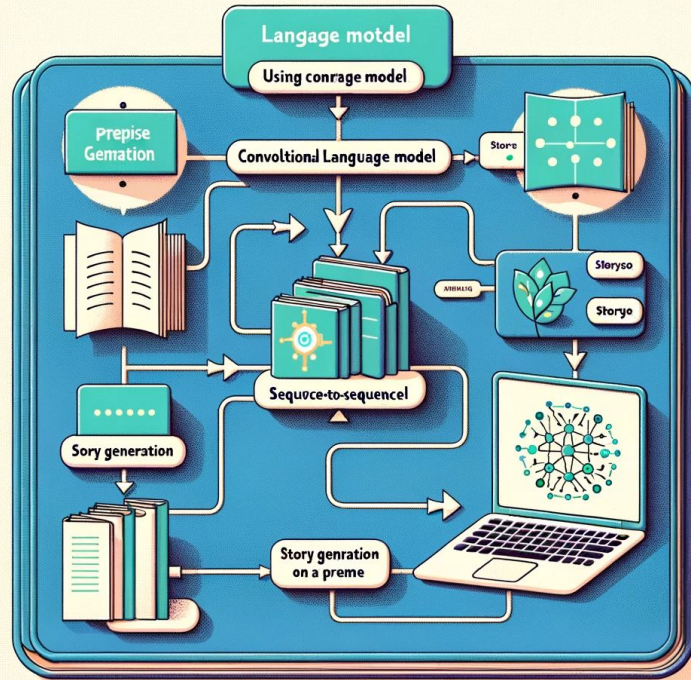


Diagram of fusion model, which learns a second seq2seq model to improve a pre- trained model. The separate hidden states are combined after gating through concatenation.

Hierarchical Story Generation Process

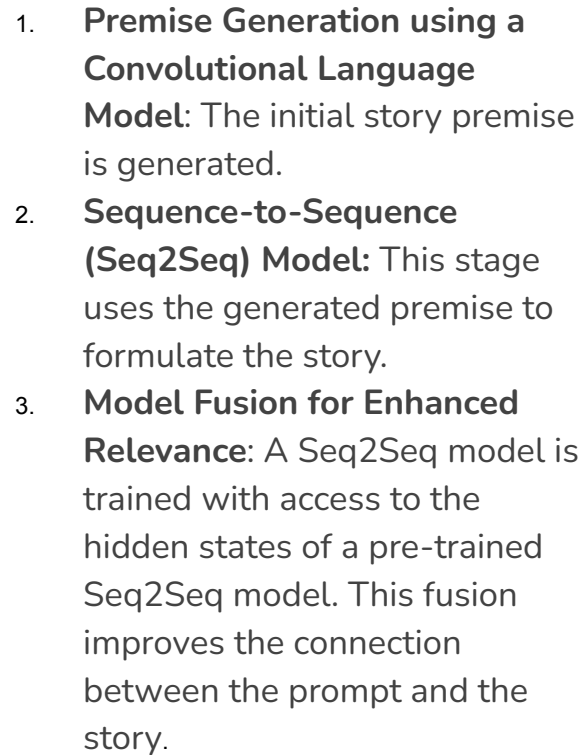


Premise Generation Using Convolutional Language Model:

- The process begins with a language model, which generates the initial premise or suggestion for the story.
- This stage uses a convolutional language model to create a basic outline or concept that forms the foundation of the story.

Story Generation Based on Premise with Sequence-to-Sequence Model:

- After the premise is generated, the next stage involves a sequence-to-sequence (seq2seq) model..
- The premise serves as a guide, helping the story remain consistent and structured beyond just individual sentences or phrases.





Automatic Evaluation and Results

Perplexity: used to evaluate the quality of language models, and it reflects how fluently the model can produce the correct next word given the preceding words

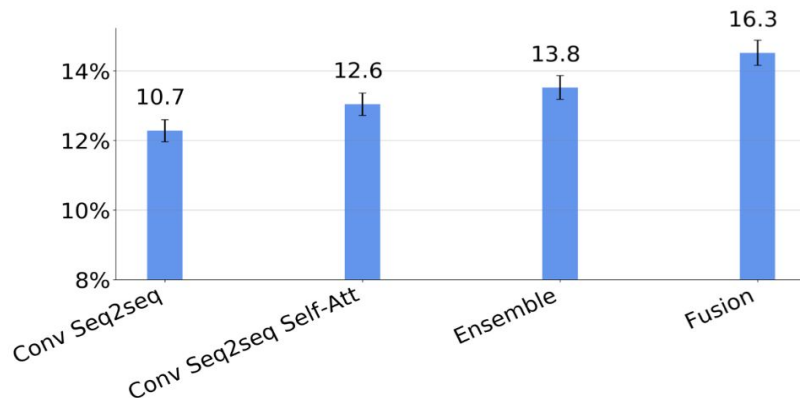
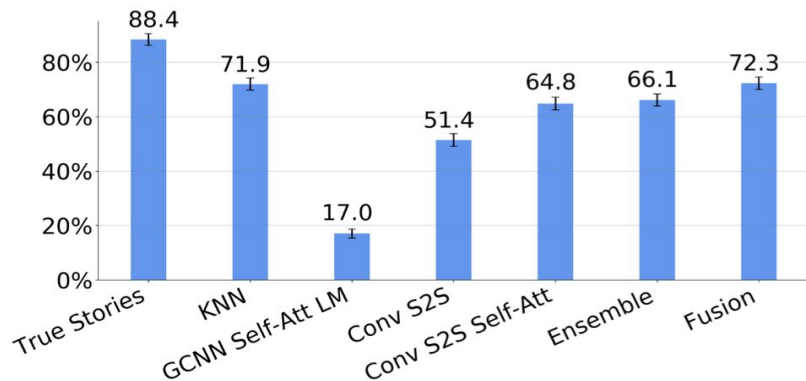
Model	# Parameters (mil)	Valid Perplexity	Test Perplexity
GCNN LM	123.4	54.50	54.79
GCNN + self-attention LM	126.4	51.84	51.18
LSTM seq2seq	110.3	46.83	46.79
Conv seq2seq	113.0	45.27	45.54
Conv seq2seq + self-attention	134.7	37.37	37.94
Ensemble: Conv seq2seq + self-attention	270.3	36.63	36.93
Fusion: Conv seq2seq + self-attention	255.4	36.08	36.56

Table 3: Perplexity on WRITINGPROMPTS. We dramatically improve over standard seq2seq models.



Human Evaluation and Results

- People find that the fusion model significantly improves the link between the prompt and generated stories





Application

Notebook



References

Max Roser (2022) - “The brief history of artificial intelligence: The world has changed fast – what might be next?” Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/brief-history-of-ai' [Online Resource].

Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In 23rd International Conference on Intelligent User Interfaces (IUI '18). Association for Computing Machinery, New York, NY, USA, 329–340. <https://doi.org/10.1145/3172944.3172983>

FAN, Angela. LEWIS, Mike. DAUPHIN, Yann. Hierarchical Neural Story Generation. Hierarchical Neural Story Generation (aclanthology.org). Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers), pages 889–898 Melbourne, Australia, July 15 - 20, 2018