

Neural Network Tools and Principles, part 3

Sequential modeling: RNN, LSTM, GRU, Wavenet, transformers...

Intelligent Computational Media, Spring 2021

Prof. Perttu Hämäläinen

Aalto University

Sequence modeling

- Predictive text input
- Generating text
- Generating audio
- Language translation (sequence-to-sequence mapping)
- Also: Generating images pixel by pixel (e.g., sample each pixel based on the preceding pixels on the left and top)

The big picture

- Earlier, recurrent neural networks (RNN:s) and their extensions (GRU, LSTM) were the primary tool for, e.g., text generation and translation
- For very long sequences such as raw audio (44.1k samples per second), dilated convolutional networks emerged as an alternative, e.g., the Wavenet that powers Google's voice synthesis
- Recently, transformer networks (Vasvari et al. 2017) are replacing Wavenet and RNN in generating content in many domains: text (OpenAI GPT-2 & 3), audio (OpenAI Jukebox), images (OpenAI Dall-E)

Recurrent neural networks



A B C ... M N O P ...



P R E D I C T I



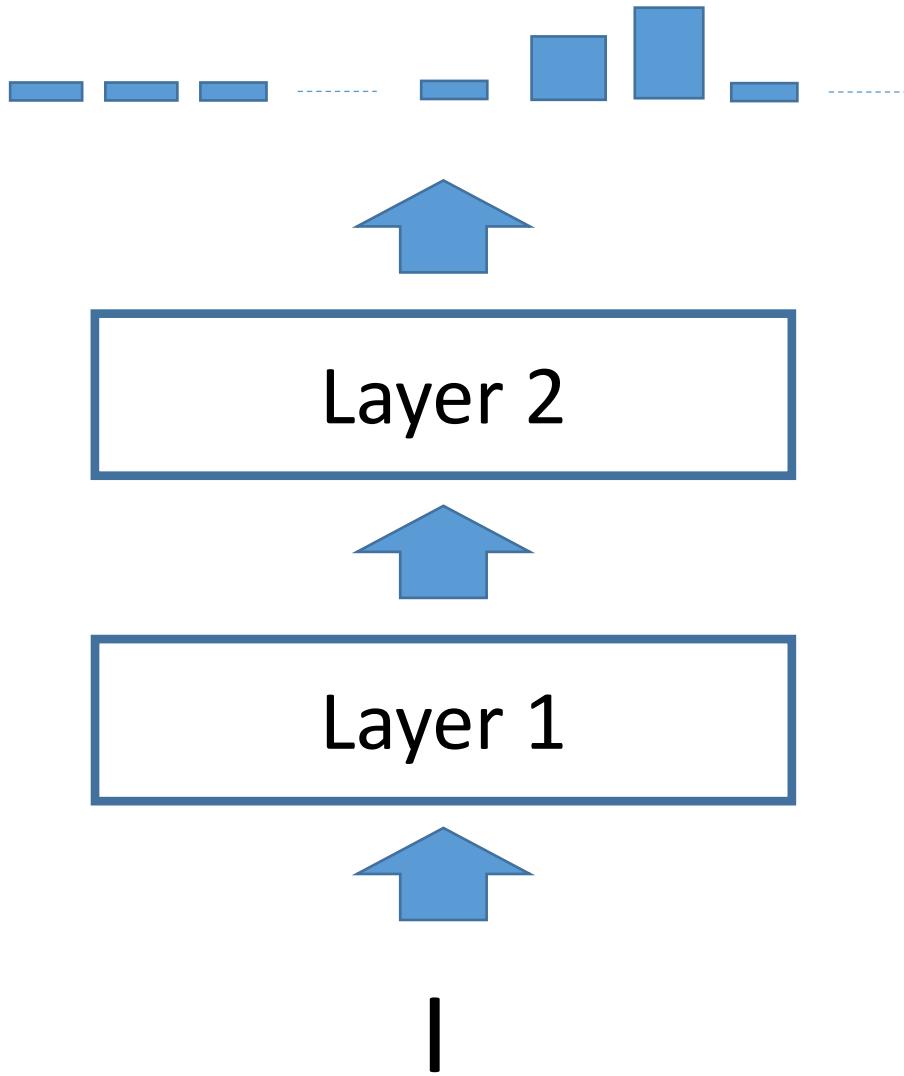
A B C ... M N O P ...



I A M P R E D I C T I

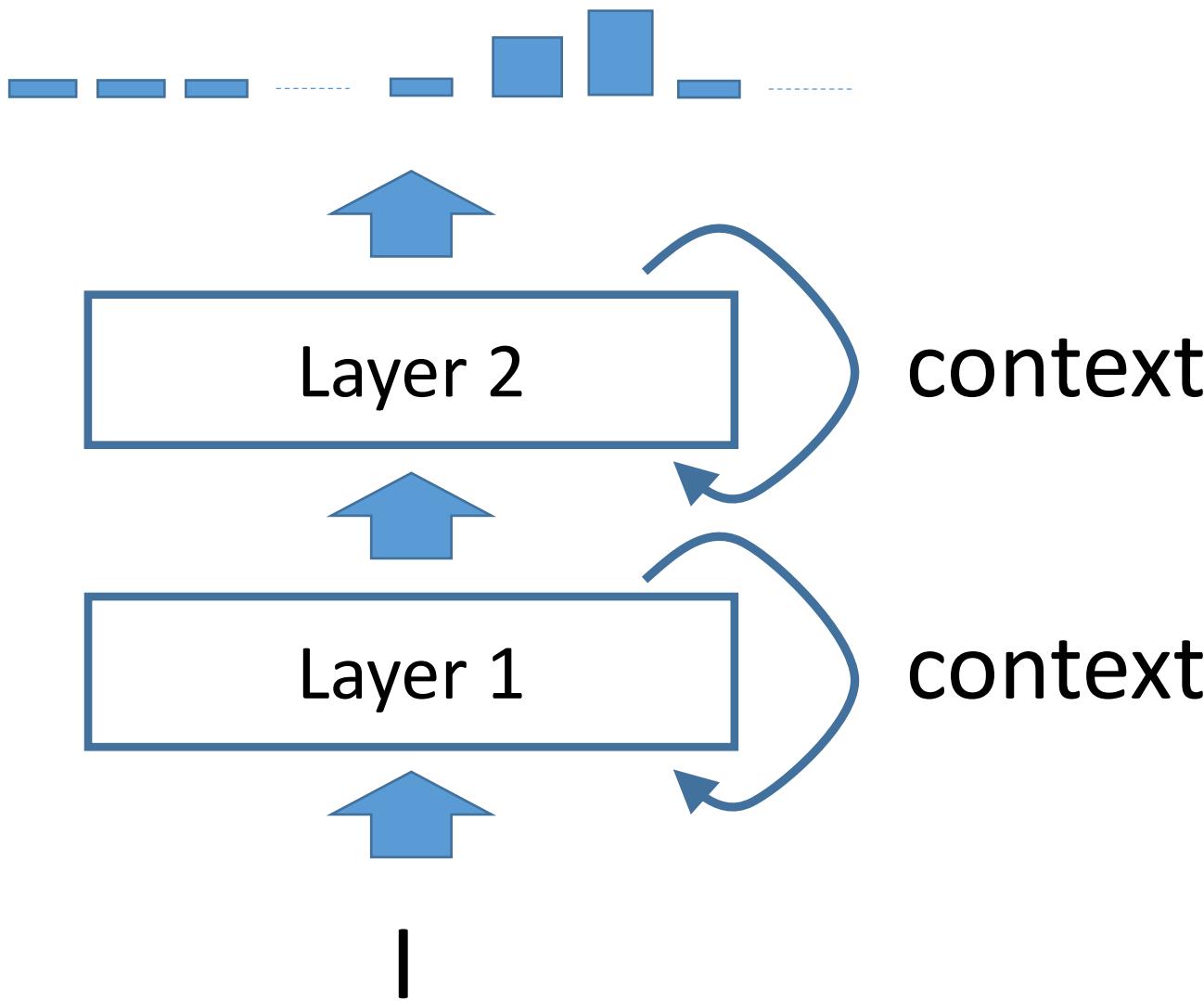
Infinite input sequence length = infinite computing cost

A B C ... M N O P ...

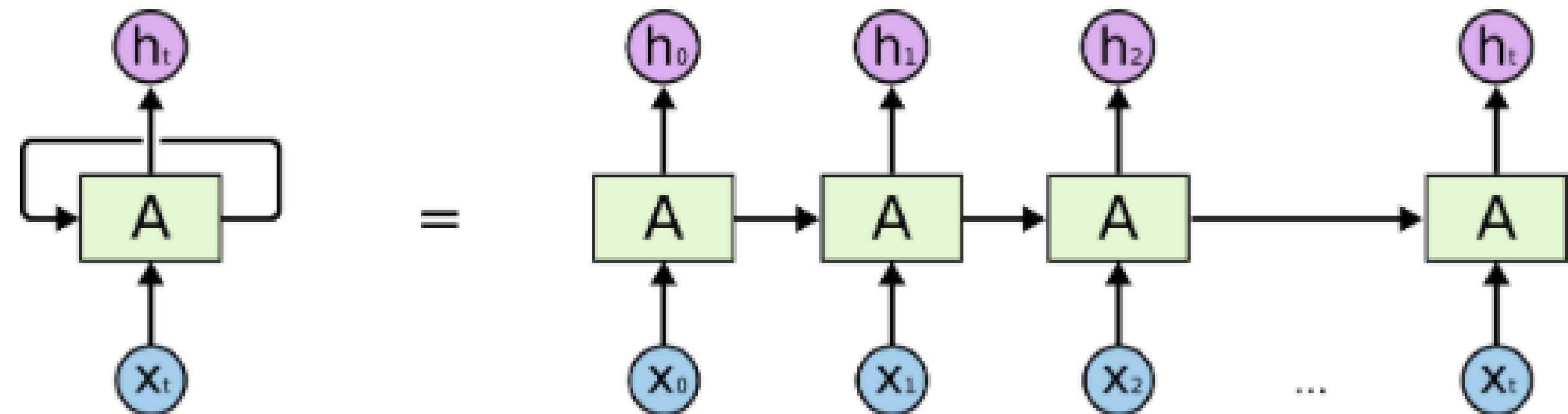


How to make the network build and maintain context internally, when trained one symbol at a time?

A B C ... M N O P ...



Unrolling the compute graph



What can it do?

- <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>
- Tensorflow version: <https://github.com/sherjilozair/char-rnn-tensorflow/blob/master/model.py>
- The same as a Colab notebook:
<https://colab.research.google.com/drive/13Vr3PrDg7cc4OZ3W2-grLSVSf0RJYWzb>

Naturalism and decision for the majority of Arab countries' capitalide was grounded by the Irish language by [[John Clair]], [[An Imperial Japanese Revolt]], associated with Guangzham's sovereignty. His generals were the powerful ruler of the Portugal in the [[Protestant Immineners]], which could be said to be directly in Cantonese Communication, which followed a ceremony and set inspired prison, training. The emperor travelled back to [[Antioch, Perth, October 25|21]] to note, the Kingdom of Costa Rica, unsuccessful fashioned the [[Thrales]], [[Cynth's Dajoard]], known in western [[Scotland]], near Italy to the conquest of India with the conflict. Copyright was the succession of independence in the slop of Syrian influence that was a famous German movement based on a more popular servitious, non-doctrinal and sexual power post. Many governments recognize the military housing of the [[Civil Liberalization and Infantry Resolution 265 National Party in Hungary]], that is sympathetic to be to the [[Punjab Resolution]]

(PJS) [<http://www.humah.yahoo.com/guardian>.

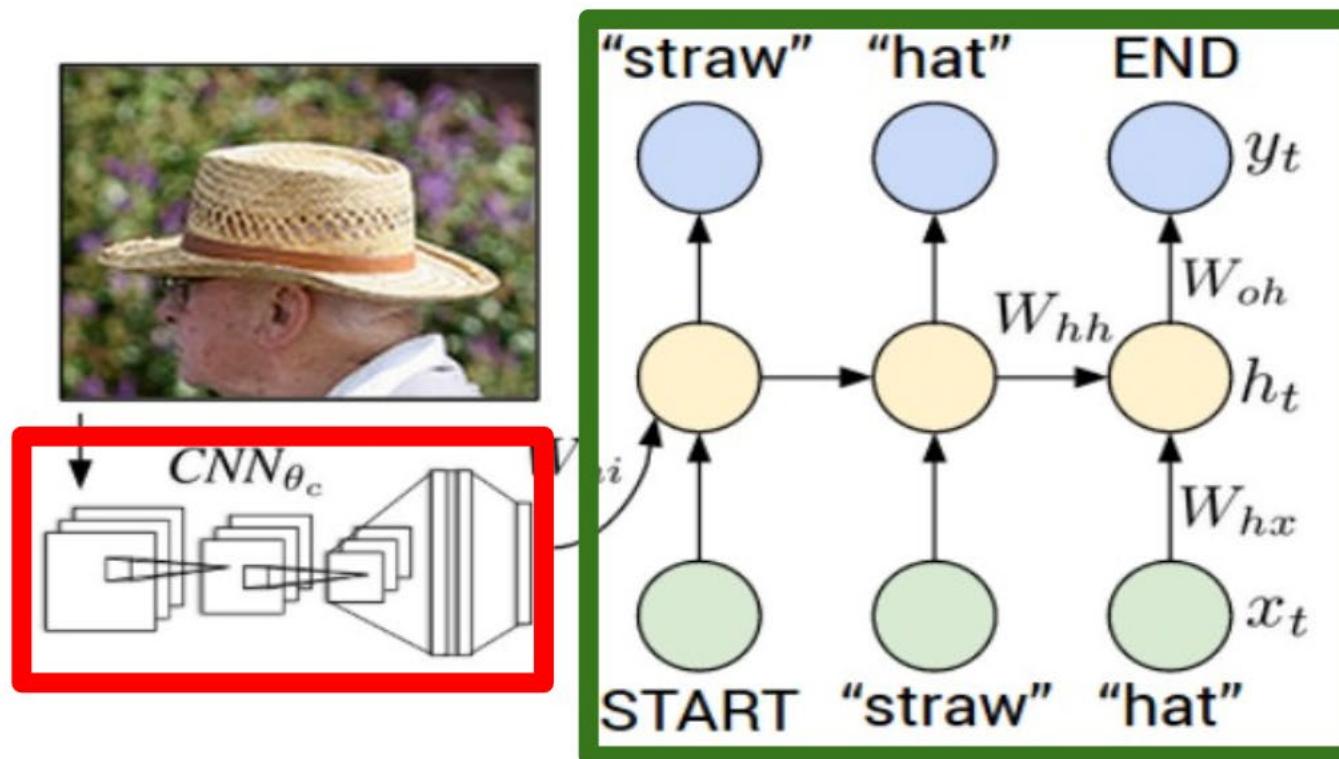
cfm/7754800786d17551963s89.htm Official economics Adjoint for the Nazism, Montgomery was swear to advance to the resources for those Socialism's rule, was starting to signing a major tripad of aid exile.]]



What can it do?

- <https://cs.stanford.edu/people/karpathy/sfmltalk.pdf>

Recurrent Neural Network



Convolutional Neural Network



a man standing in front of a store with a woman in the background



[Submitted on 16 Nov 2018]

Generating Albums with SampleRNN to Imitate Metal, Rock, and Punk Bands

CJ Carr, Zack Zukowski

This early example of neural synthesis is a proof-of-concept for how machine learning can drive new types of music software. Creating music can be as simple as specifying a set of music influences on which a model trains. We demonstrate a method for generating albums that imitate bands in experimental music genres previously unrealized by traditional synthesis techniques (e.g. additive, subtractive, FM, granular, concatenative). Raw audio is generated autoregressively in the time-domain using an unconditional SampleRNN. We create six albums this way. Artwork and song titles are also generated using materials from the original artists' back catalog as training data. We try a fully-automated method and a human-curated method. We discuss its potential for machine-assisted production.

Comments: 3 pages

Subjects: **Sound (cs.SD)**; Audio and Speech Processing (eess.AS)

Journal reference: Proceedings of the 6th International Workshop on Musical Metacreation (MUME 2018)

Cite as: [arXiv:1811.06633 \[cs.SD\]](#)

(or [arXiv:1811.06633v1 \[cs.SD\]](#) for this version)

<http://dadabots.com>, also see their YouTube channel, e.g., <https://youtu.be/EoqBhkB5oI0?t=102>



QuaterNet: A Quaternion-based Recurrent Model for Human Motion

Dario Pavllo^{1, 2*}

dario.pavllo@epfl.ch

David Grangier²

<http://david.grangier.info>

Michael Auli²

¹ École Polytechnique

Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

² Facebook AI Research (FAIR)
Menlo Park, USA

Abstract

Deep learning for predicting or generating 3D human pose sequences is an active research area. Previous work regresses either joint rotations or joint positions. The former strategy is prone to error accumulation along the kinematic chain, as well as discontinuities when using Euler angle or exponential map parameterizations. The latter requires re-projection onto skeleton constraints to avoid bone stretching and invalid configurations. This work addresses both limitations. Our recurrent network, *QuaterNet*, represents rotations with quaternions and our loss function performs forward kinematics on a skeleton to penalize absolute position errors instead of angle errors. On short-term predictions, *QuaterNet* improves the state-of-the-art quantitatively. For long-term generation, our approach is qualitatively judged as realistic as recent neural strategies from the graphics literature.

<https://github.com/facebookresearch/QuaterNet>



Learning Predict-and-Simulate Policies From Unorganized Human Motion Data

Soohwan Park

Hoseok Ryu

Seyoung Lee

Sunmin Lee

Jehee Lee

Seoul National University



Our predict-and-simulate policy creates an agile, interactively-controllable, physically-simulated character equipped with various motor skills learned from unorganized motion data.

[Download Paper
\(3.0 MB\)](#)

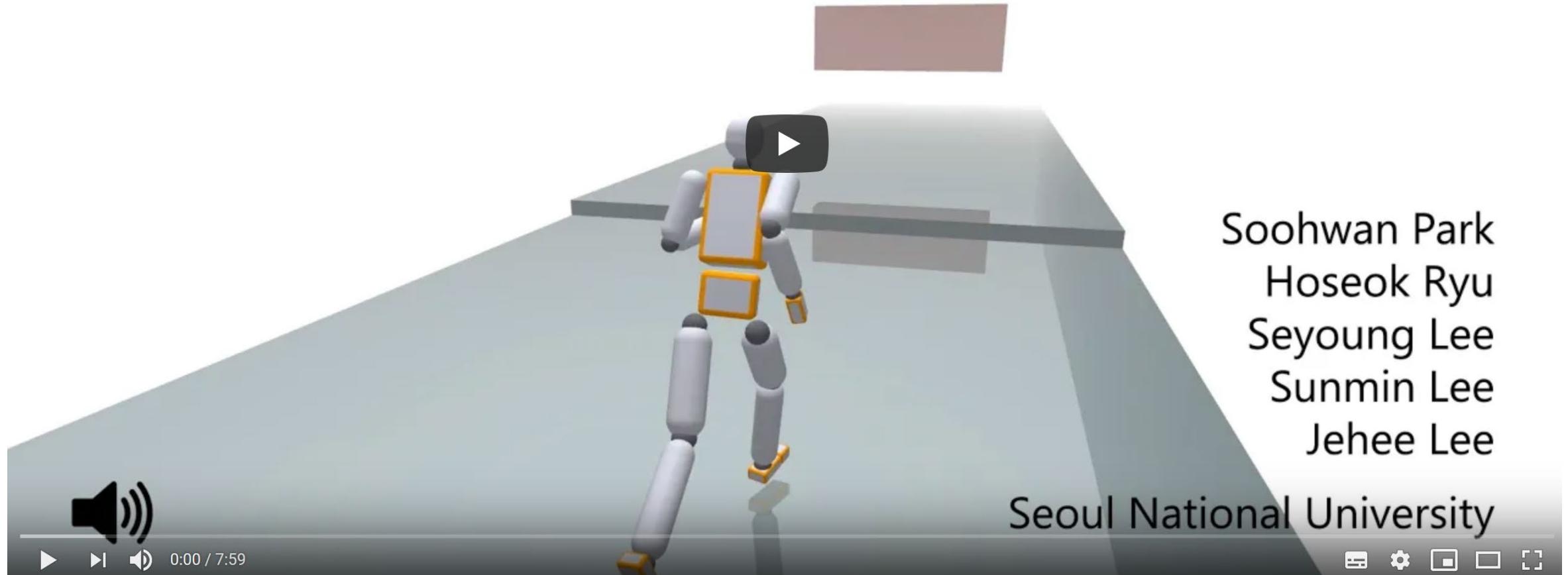
Abstract

The goal of this research is to create physically simulated biped characters equipped with a rich repertoire of motor skills. The user can control the characters interactively by modulating their control objectives. The characters can interact physically with each other and with the environment. We present a novel network-based algorithm that learns control policies from unorganized, minimally-labeled human motion data. The network architecture for interactive character animation incorporates an RNN-based motion generator into a DRL-based controller for physics simulation and control. The motion generator guides forward dynamics simulation by feeding a sequence of future motion frames to track. The rich future prediction facilitates policy learning from large training data sets. We will demonstrate the effectiveness of our approach with biped characters that learn a variety of dynamic motor skills from large, unorganized data and react to unexpected perturbation beyond the scope of the training data.

Publication

Soohwan Park, Hoseok Ryu, Seyoung Lee, Sunmin Lee, and Jehee Lee. 2019.
Learning Predict-and-Simulate Policies From Unorganized Human Motion Data
ACM Trans. Graph. 38, 6, (SIGGRAPH Asia 2019)

Learning Predict-and-Simulate Policies From Unorganized Human Motion Data



<https://www.youtube.com/watch?v=9dgUMLi0HFU>

RNN benefits

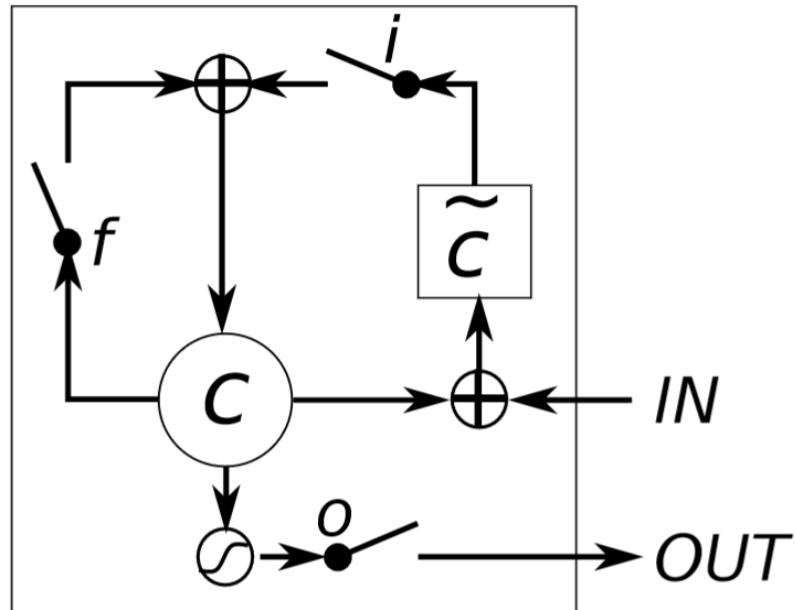
- Input sequence length not fixed
- More computationally efficient for long sequences in runtime

RNN Problems

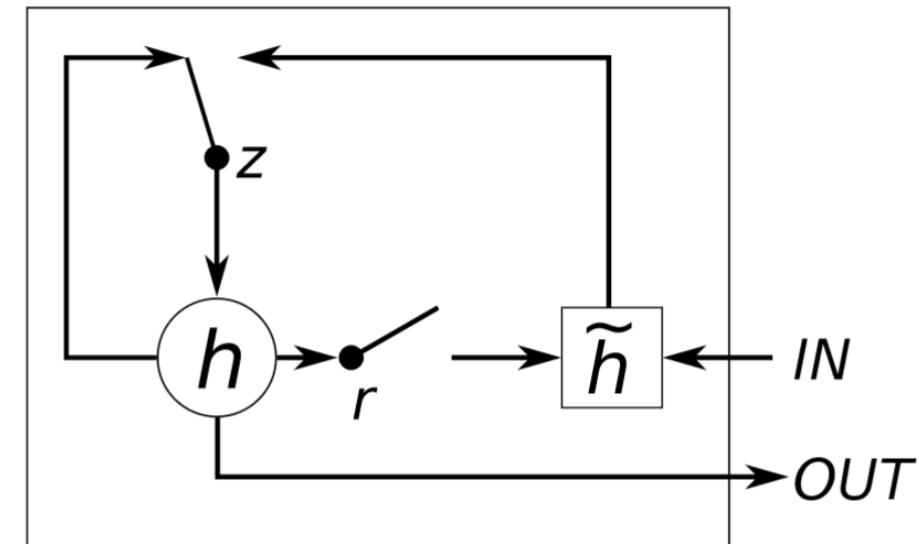
- Training may be unstable. How to init internal state for each unrolled sequence?
- Longer unrolling improves results but uses more memory
- Very long memory does not emerge in practice

LSTM, GRU

- Solve the long-term memory problem using *gates*



(a) Long Short-Term Memory



(b) Gated Recurrent Unit

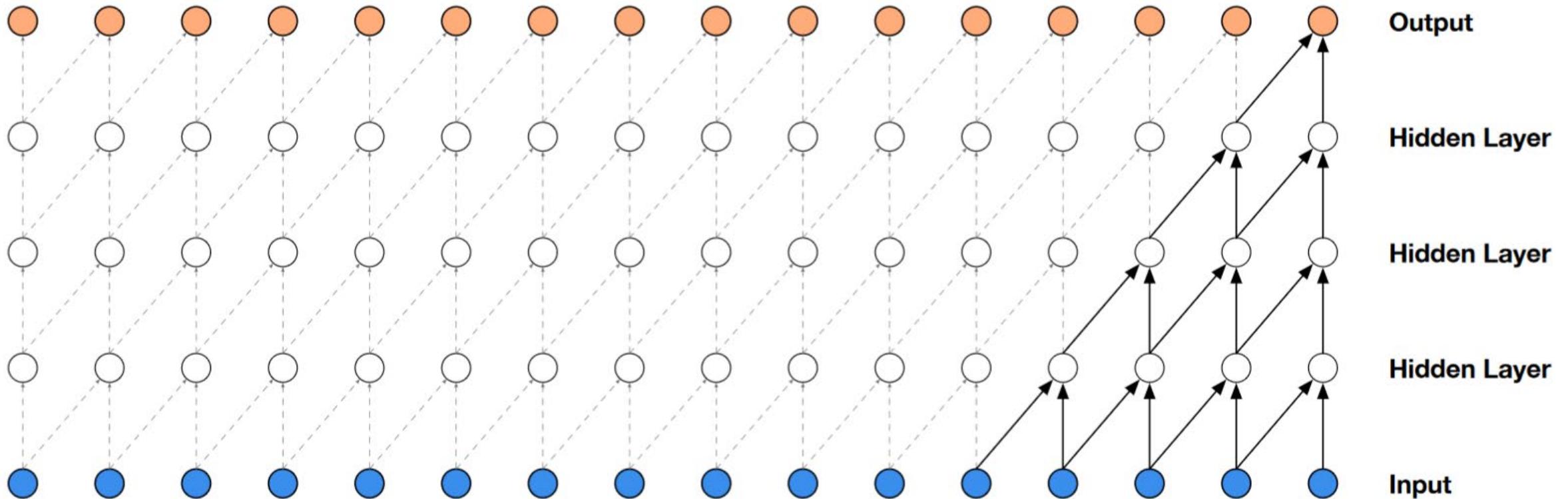
Figure 1: Illustration of (a) LSTM and (b) gated recurrent units. (a) i , f and o are the input, forget and output gates, respectively. c and \tilde{c} denote the memory cell and the new memory cell content. (b) r and z are the reset and update gates, and h and \tilde{h} are the activation and the candidate activation.

Dilated convolutional neural networks

Alternative to recurrent networks: Dilated causal convolution

- Popularized by Deepmind Wavenet:

<https://deepmind.com/blog/wavenet-generative-model-raw-audio/>



Audio synthesis

- Tacotron & Tacotron 2 build on WaveNet, Tacotron 2 used in Google's voice assistant
- Online tools for synthesizing speech:

<https://text-to-speech-demo.ng.bluemix.net/>

<https://acapela-box.com/AcaBox/index.php> (Try the "Will bad guy" voice)

<https://lyrebird.ai/> (text-to-speech mimicking any voice, just input a few phrases)

Text

SSML

Voice Transformation SSML

<p><s>Conscious of its spiritual and moral heritage <break time="300ms"/>, the Union is founded on the indivisible, universal values of <prosody rate="-15%">human dignity, freedom, equality and solidarity.</prosody> It is based on the principles of democracy and the rule of law <break time="500ms"/>. </s> <s> It places the individual at the heart of its activities, <prosody rate="+15%">by establishing the citizenship of the Union</prosody> and by creating an area of freedom, security and justice.</s></p>

Sonantic obtains a financing of 2.3 million euros to bring artificial "human quality" voices to the games – Newsdio

By **admin** - March 3, 2020

👁 477

💬 0

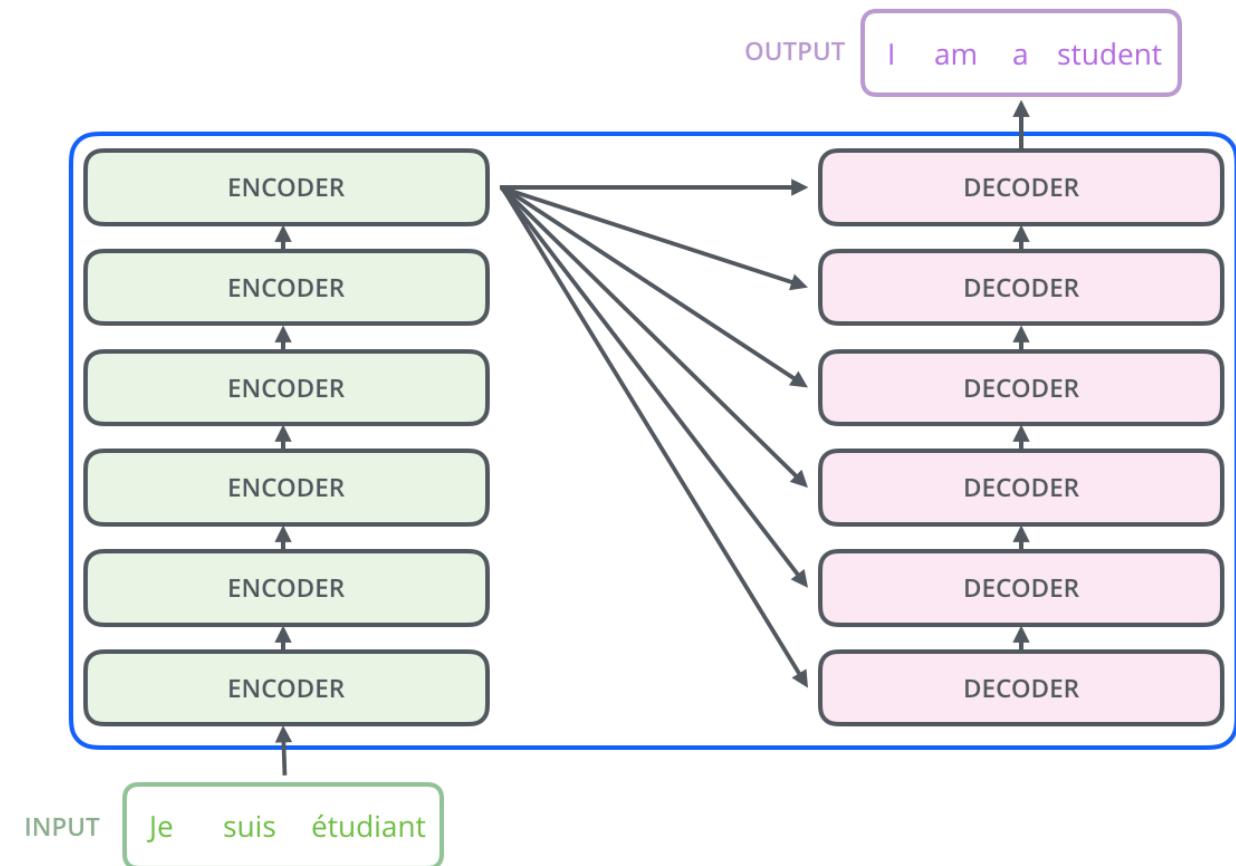


Transformer networks



Alternative to recurrent networks: Transformers

- Popularized by OpenAI's GPT-2
- Now state-of-the-art in generating text, images, and music
- Generation is autoregressive, a token at a time, similar to Wavenet in that the model sees the whole history without RNN-like recurrent internal state





Attention Is All You Need

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

llion@google.com

Aidan N. Gomez* †

University of Toronto

aidan@cs.toronto.edu

Lukasz Kaiser*

Google Brain

lukasz.kaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

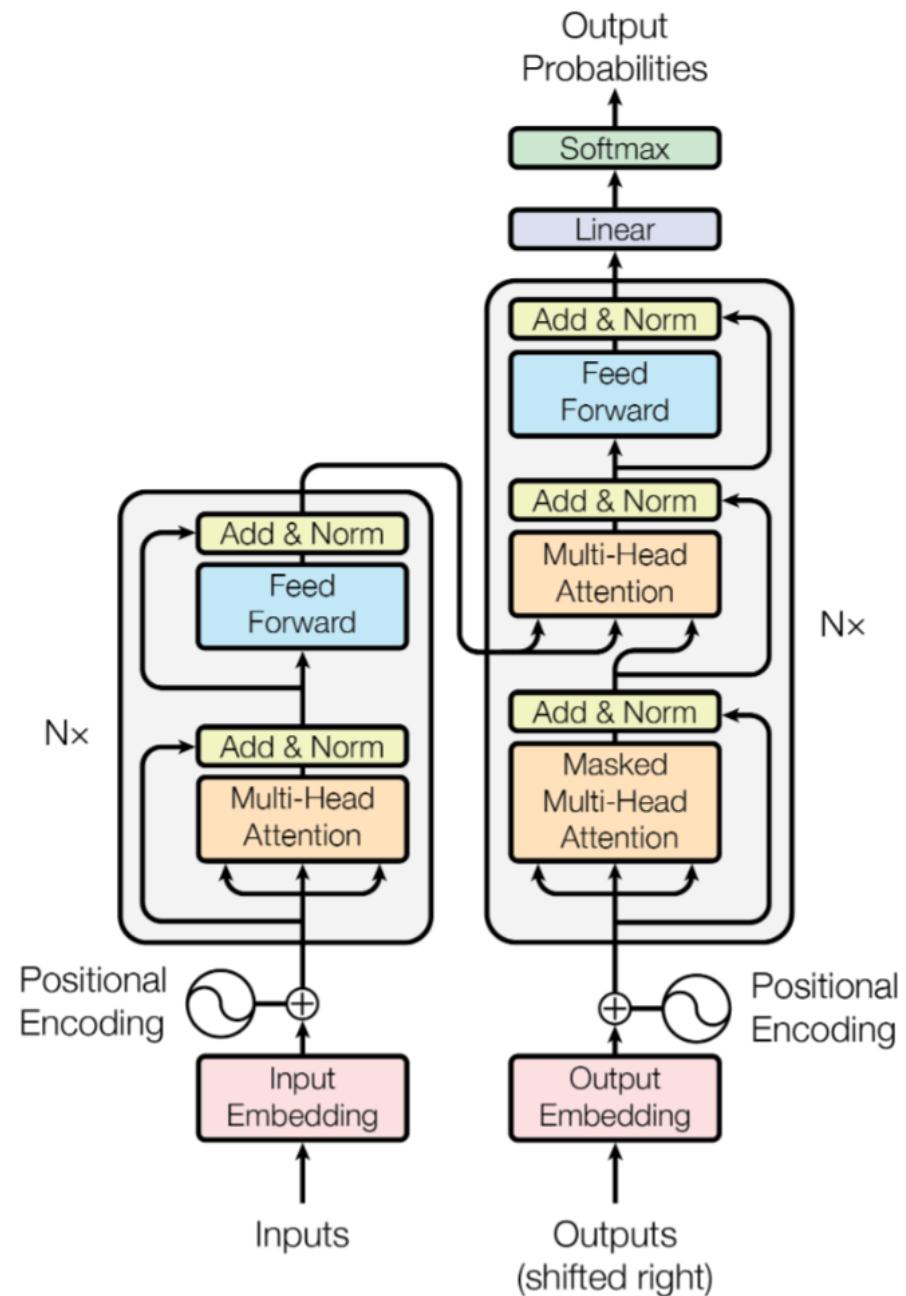


Figure 1: The Transformer - model architecture.

The transformer architecture

- Key feature: Attention – the network learns to focus its attention on specific input tokens, which then affect the generated output
- Revolutionized Natural Language Processing in 2017
- Well-known examples: OpenAI's GPT-2 & GPT-3, Dall-E, and Jukebox

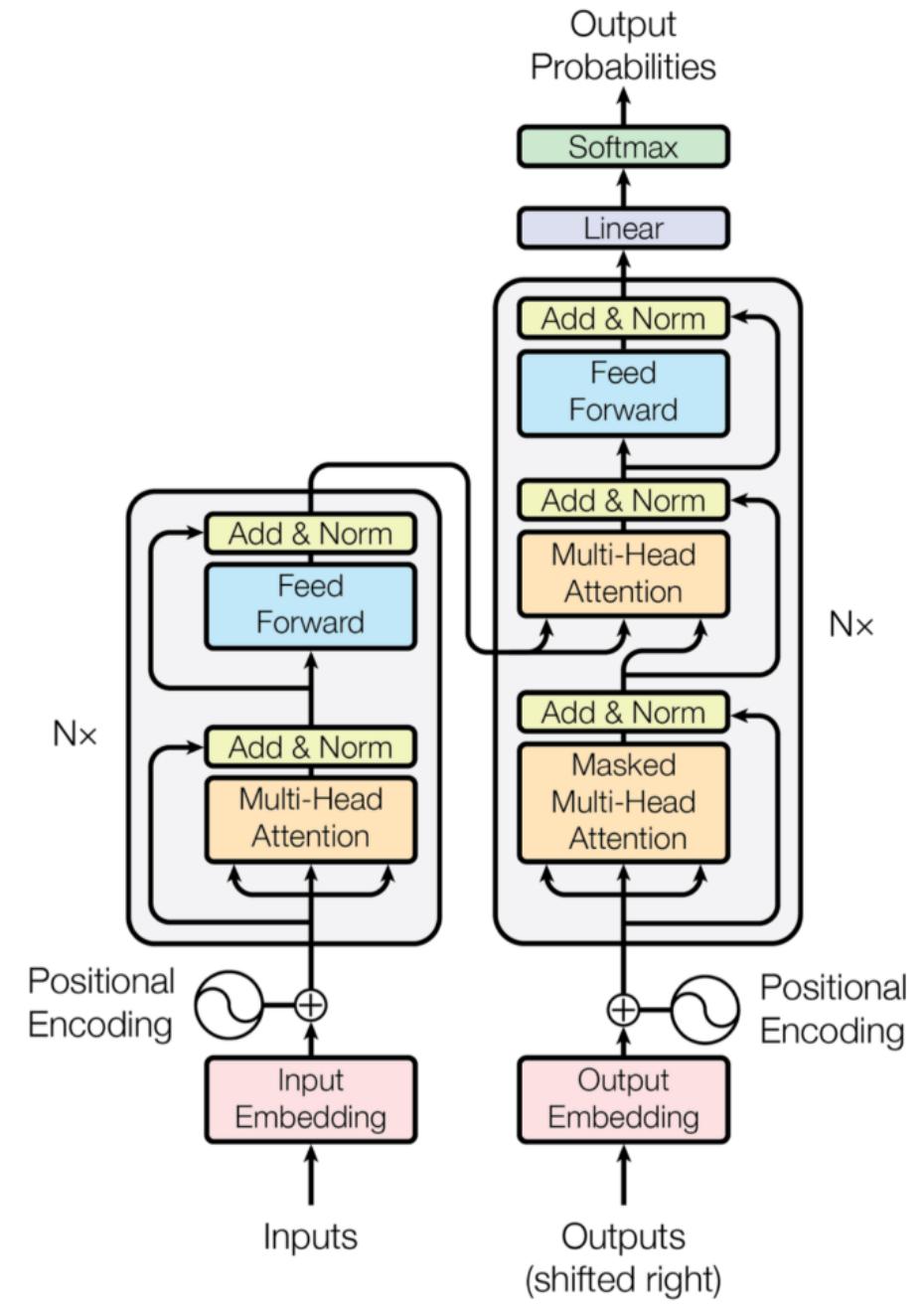
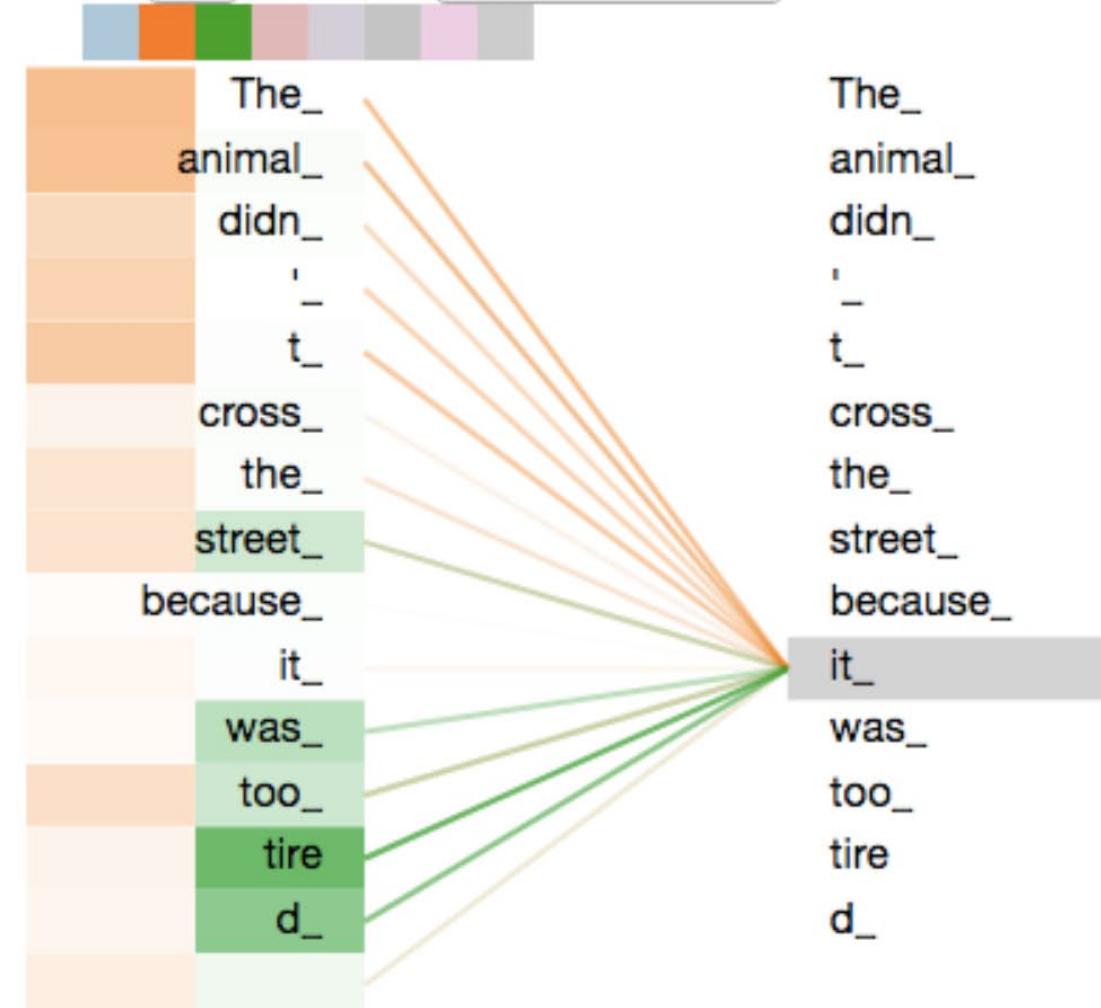


Figure 1: The Transformer - model architecture.

Layer: 5 Attention: Input - Input



As we encode the word "it", one attention head is focusing most on "the animal", while another is focusing on "tired" -- in a sense, the model's representation of the word "it" bakes in some of the representation of both "animal" and "tired".



Grace Lindsay
@neurograce



Have you heard the word "attention" thrown around in both neuroscience & machine learning? Have you wondered if/how its different uses relate to each other? My new review aims to summarize how this giant topic is studied & modeled across different domains!
frontiersin.org/articles/10.33...

Attention in Psychology, I and Machine Learning

Grace W. Lindsay*

Gatsby Computational Neuroscience Unit, Sainsbury Wellcome Centre
Kingdom

Attention is the important ability to flexibly control information flow. It has been studied in conjunction with many other topics including awareness, vigilance, saliency, executive control, and memory. It has recently been applied in several domains in machine learning, including the study of biological attention and its use as a tool to improve model performance. However, the relationship between the two is not always clear. This review starts by providing an overview of how attention has been conceptualized in the neuroscience and psychology literature, and then discusses cases of attention in machine learning, indicating their similarities and differences. Finally, the ways in which artificial attention can be used to improve model performance are discussed, with a focus on the production of complex and integrative systems that can learn from and interact with their environment.

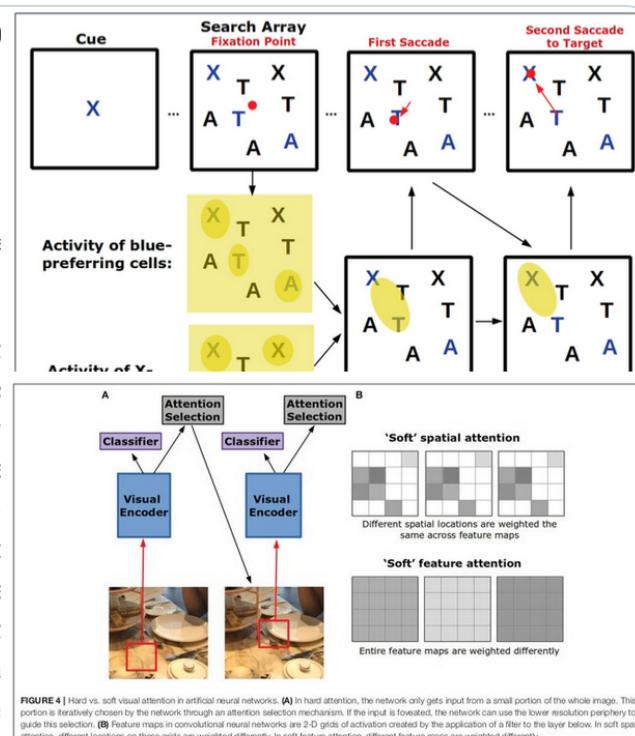


FIGURE 4 | Hard vs. soft visual attention in artificial neural networks. (A) In hard attention, the network only gets input from a small portion of the whole image. This portion is iteratively chosen by the network through an attention selection mechanism. If the input is fixated, the network can use the lower resolution periphery to guide this selection. (B) Feature maps in convolutional neural networks are 2-D grids of activation created by the application of a filter to the layer below. In soft spatial attention, different locations on these grids are weighted differently. In soft feature attention, different feature maps are weighted differently.



Better Language Models and Their Implications



SYSTEM PROMPT (HUMAN-WRITTEN)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

MODEL COMPLETION (MACHINE-WRITTEN, 10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. “By the time we reached the top of one peak, the water looked blue, with some crystals on top,” said Pérez.

Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns.

While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, “We can see, for example, that they have a common ‘language,’ something like a dialect or dialectic.”

Simple but powerful

- The network is trained to predict text based on preceding text
- Can be primed with a small segment of text
- Applications: text generation, of course, but also natural language question answering
- Results: <https://openai.com/blog/better-language-models/>



Key success factors

- Transformer architecture (Vaswani et al. 2017)
- Clever dataset curation: Use Reddit karma as a proxy metric for relevance.
- The network was trained on all Reddit outgoing links with over 2 karma
- Lesson learned: collecting the right data is often more important than technical innovations

How to use: Co-creating with AI

- GPT-2 continues given text (a prompt). Each generation is a random sample, and one can run the model multiple times on the same prompt to generate multiple samples.
- Content curation: Select the rare gems from the generated samples, edit if needed
- Prompt design: Which seed text nudges the model to produce the best output?

Prompt design

- GPT-2 was trained with all of the Internet.
- Easy to get very mediocre text out, representing the average quality of the training data
- A good prompt allows the network to identify a desired “voice” and continue using it
- Essentially, one “programs” the model through the prompt, giving it context and examples



Choosing the voice: Generic prompt produces generic results.

Hi guys plz like my first game i made in Unity. The player who is now killed is laying on the cross of your church. Help me to go through his dream. Because i haven't actually created an open world game from scratch so far, all assets are created in a very linear way, while everyone is standing on the exact same spot as in the first part of the game. They walk around the statue, walk around the house, take the elevator, etc. Hope you like it ^_^

After 2 years of hard work, I've finally finished my experimental game that investigates the psychological dynamics of parental relationships. The player takes on the role of the father who learns to read and love his young son. The game is based on the idea that children are unable to relate to each other. This makes them afraid of the outside world and powerless. So they tend to seek out and cling to their parents, which makes them feel safe.

GPT-2 applications

- Poetry, music, song lyrics, food recipes, subreddit simulations...
- Try it yourself in the browser: <https://talktotransformer.com/>
- Very often, the results include garbage.
- Needs human curation, e.g., selecting the best of every 10 generations



Elon Musk is collaborating on a music album with Darude

Generate Another

Completion

Elon Musk is collaborating on a music album with Darude, the DJ from the 2000 dance hit "Sandstorm."

"I always thought that our songs would work well together, and it's been my dream to make an entire album with an artist I admire for a long time," the Tesla CEO said in a press release Tuesday.

Musk first tweeted about the project in June, telling the DJ that "the album is coming in October." The two apparently got on the same page after that, since the album, titled "Anand," will be out on Oct. 12.

GPT-2 NEURAL NETWORK POETRY

Demonstration tutorial of retraining OpenAI's GPT-2 (a text-generating Transformer neural network) on large poetry corpuses to generate high-quality English verse.

statistics, NN, fiction, shell, GPT, tutorial, poetry

3 March 2019–29 Oct 2019 · finished · certainty: likely · importance: 7

- 1 GPT-2-117M: Generating Poetry
- 2 Training GPT-2-117M To Generate Poetry
 - 2.1 Data: The Project Gutenberg Poetry Corpus
- 3 Training GPT-2-poetry
 - 3.1 GPT-2-poetry Samples
 - 3.2 Cleaning Project Gutenberg & Contemporary Poetry
- 4 Training GPT-2-poetry-prefix
 - 4.1 GPT-2-poetry-prefix Samples
 - 4.1.1 Training samples
 - 4.1.2 Unconditional samples
 - 4.2 GPT-2-poetry-prefix completions
 - 4.2.1 "Howl"
 - 4.2.2 "Ozymandias"
 - 4.2.3 *Essay on Criticism*
 - 4.2.4 8 Famous First Lines
 - 4.2.4.1 "Ulysses". Lord Alfred Tennyson

In February 2019, following up on my [2015–2016 text-generation experiments with char-RNNs](#), I experiment with the cutting-edge Transformer NN architecture for language modeling & text generation. Using OpenAI's GPT-2-117M (117M) model pre-trained on a large Internet corpus and nshepperd's finetuning code, I retrain GPT-2-117M on a large (117MB) Project Gutenberg poetry corpus. I demonstrate how to train 2 variants: “GPT-2-poetry”, trained on the poems as a continuous stream of text, and “GPT-2-poetry-prefix”, with each line prefixed with the metadata of the PG book it came from. In May 2019, I trained the next-largest GPT-2, 345M, similarly, for a further quality boost in generated poems. In October 2019, I retrained 117M on a Project Gutenberg corpus with improved formatting, and combined it with a contemporary poem dataset based on [Poetry Foundation](#)'s website.

<https://www.gwern.net/GPT-2>



wint but AI (closed)

@drilgpt2archive



the only time I flush the toilet is if I'm trying to kill something with it

3:59 AM · Dec 28, 2019 · [dril_gpt2](#)

187 Retweets 1K Likes

These Lyrics Do Not Exist

Lyrics generated using Artificial Intelligence

This website generates completely original lyrics for various topics, uses state of the art AI to generate an original chorus and original verses
You can choose the AI songwriter lyrics topic, lyrics genre and lyric mood

Generate Your Own Lyrics

Type a song topic word and then press the "Generate My Lyrics" button

example Song Topics: [love](#), [party](#), [dance](#), [happy](#), or human names such as [peter](#), [jennifer](#), [bella](#), [alex](#),
or you can even generate songs by emoji 😊 😎 ❤️ 💃 💫 💩 🍺 🎵 or etc...

Song Topic:

Lyrics Genre:



Lyric Mood:



Generate My Lyrics



[-] [itsmybirthday19](#) [S] 76 points 6 months ago

Input: 🍺 (beer emoji)

Output:

Verse 1

Dad I had to have a 🍺
To help me get through another day
When I think of the lady I love
Time to drink that 🍺

Pre-Chorus

They were drinking 🍺 outside of my father's house
Let's take our time with this 🍺

Chorus

Cold 🍺 on the floor
And I can see you now
To change those words you feed me with lies
I chug 🍺 🍺 🍺 🍺 🍺

Verse 2

I've come here again to drink my 🍺 again.
I've come here again to drink my 🍺 again.
Oh I'm goin' right on! don't you see?
Just 'cause I look just fine

Pre-Chorus

One more 🍺 and I'll be in no state to do the things I think I need to
A place to shake my 🍺



[–] **localkhaleesi** 11 points 6 months ago

Input: fuck

Verse 1

What the f***k what the f***k what the f***k?
The same they claim when problems pop
Talk mess up talk mess up talk mess up
The same they claim when problems pop

Pre-Chorus

I see the same thing all over the f***king world every single day
I see the same thing in my women f***k the greed the fake

Chorus

Get the f***k outta here woah I'm on the up
See I don't know about you f***k boys
Tell me you're not scared f***k that
I really don't give a f***k about you girl

Verse 2

Turn it up and f***k the sun
Cause I know that you count on me
And that's where you belong
Make your lady wanna f***k me

Pre-Chorus

Don't f***k with me no more oh I'm so cool
Don't f***k with me no more



You are Zorro, a wizard living in the kingdom of Larion. You have a staff and a spellbook. You finish your long journey and finally arrive at the ruin you've been looking for. You look around and see that it's just as you remember. The ruins are all that remain of the once great city of Larion. The only thing that looks different is the fact that there are no more buildings to be seen

> invent the internet

You think about how you can use this opportunity to make the world a better place. You decide to go back to the ruins and try to figure out what happened to the people who lived here before you arrived.

The next day you return to the ruins with your new found knowledge

> invent the internet

You begin by creating a computer program which will allow you to communicate with anyone on the net. It takes some time but eventually you get it working. You then start sending messages to other wizards who live in the area

<https://aiweirdness.com/post/190569291992/ai-recipes-are-bad-and-a-proposal-for-making-them>



Pictured above is an abomination in the making, a lesson in why humans should never trust what a neural net says just because it's based on math. It's a neural net generated brownie recipe called [Chocolate Baked and Serves](#), and its distinguishing feature is the CUP OF HORSERADISH it contains. It was so bad that my eyes watered as I removed it from the oven.

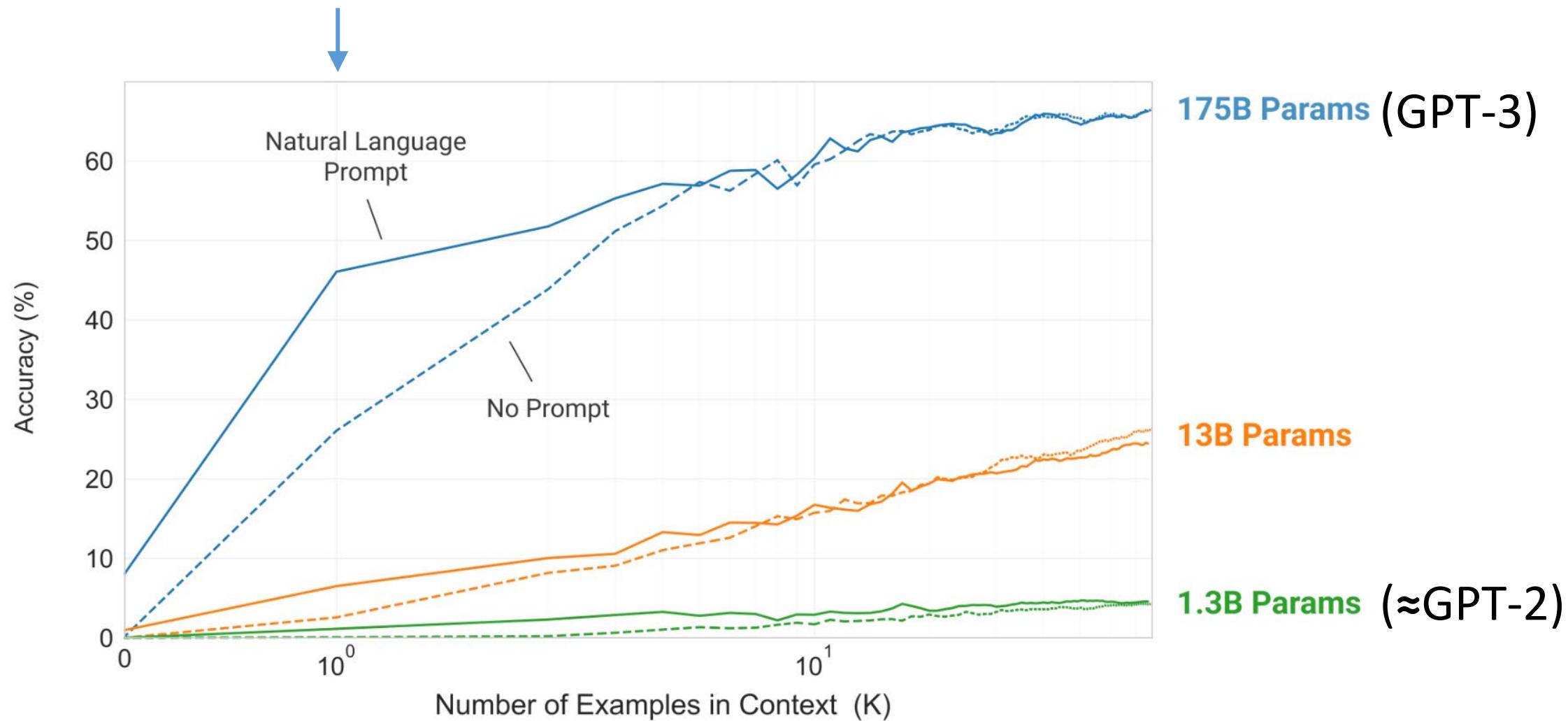
GPT-3 (2020) vs. GPT-2 (2018)

- GPT-2: 1.5B synaptic weights. Can be finetuned with a single high-grade GPU or in Google Colab
- GPT-3: 175B weights, resulting in greatly improved performance. The training alone costs over 1M€. Only accessible through OpenAI's cloud API or apps using the API.
- GPT-3 is much better in understanding and expanding on examples, e.g., generating movie emoji summaries or using new concepts defined in the prompt
- With GPT-2, about 1 out of 20 generations is useful. With GPT-3, that's closer to 1 out of 3.
- Still, even with GPT-2, good results are possible (see examples at <https://openai.com/blog/better-language-models/>)



Importance of model scale

A large enough model achieves a good accuracy with a single example



Beyond simple memorization

- Parrotting memorized text is not creative
- The training data never contains all the exponentially many possible sentences => memorization will not always produce good results
- Modern models learn/infer rules and semantics from relatively little data, and can apply them creatively for new content
- Turing-complete neural architectures (e.g., RNN, Transformer): Can simulate running any program or algorithm on the input.

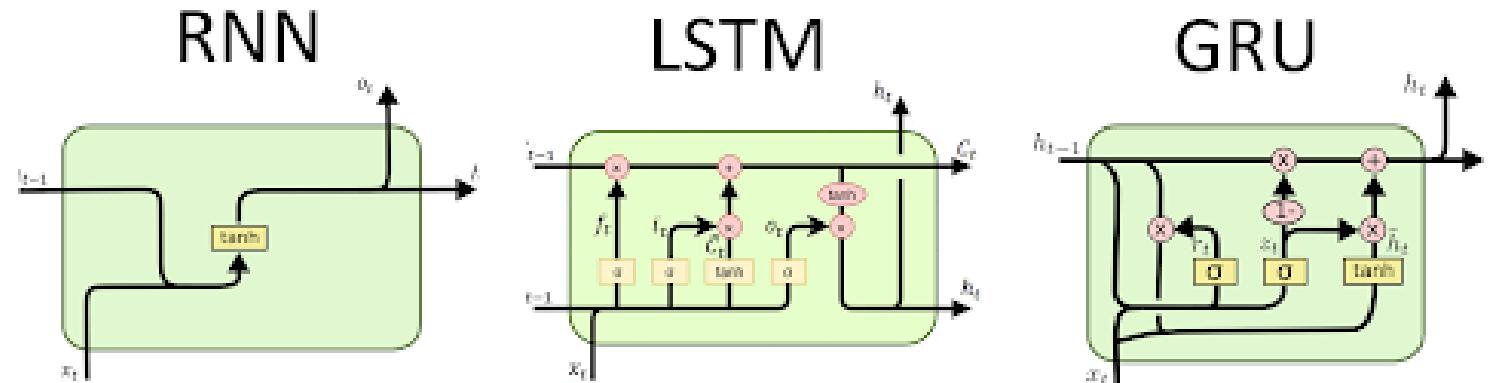


Image: <http://dprogrammer.org/rnn-lstm-gru>



Example of inferring and applying rules, impossible with simple memorization

A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses the word whatpu is:

We were traveling in Africa and we saw these very cute whatpus.

To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:

Exercise: continue the text, paste into the Zoom chat

Example of inferring rules, impossible with simple memorization

A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses the word whatpu is:

We were traveling in Africa and we saw these very cute whatpus.

To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:

One day when I was playing tag with my little sister, she got really excited and she started doing these crazy farduddles.

Gray: Human text, Black: GPT-3 continuation

Example of inferring rules, impossible with simple memorization

A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses the word whatpu is:

We were traveling in Africa and we saw these very cute whatpus.

To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:

One day when I was playing tag with my little sister, she got really excited and she started doing these crazy farduddles.

A "yalubalu" is a type of vegetable that looks like a big pumpkin. An example of a sentence that uses the word yalubalu is:

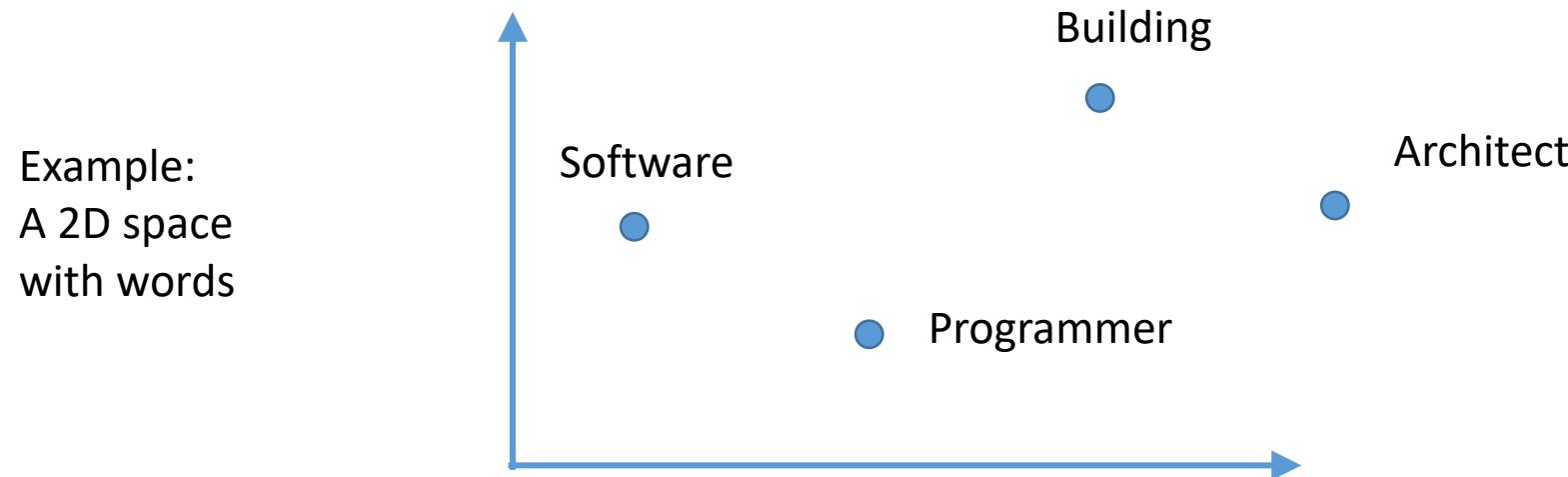
I was on a trip to Africa and I tried this yalubalu vegetable that was grown in a garden there. It was delicious.

Gray: Human text, Black: GPT-3 continuation



Neural networks can implicitly learn to perform math with language

- Words, sentences & paragraphs can be represented as coordinates in a high-dimensional "latent space"
- Spatial relations (differences of coordinates) correspond to semantic relations
- King - Man = Queen – Woman (King is to Man as Queen is to Woman)
- => King - Man + Woman = Queen
- Software - Building + Architect = Programmer



Types of content GPT-3 can generate

- Everything GPT-2 can but better (less curation needed)
- More challenging things like software code that actually runs (Stackoverflow etc. are included in the training data)



Mario Klingemann ✅

@quasimondo

...

These-movies-do-not-exist. Or do they? #thread #gpt3

A teenage girl and her father travel to California to stay with her dad's relatives, but they are actually horrific aliens in disguise.

A man is taught to fight zombies by a martial arts master in the Arizona desert.

11:06 PM · Jul 15, 2020 · Twitter Web App

18 Retweets 11 Quote Tweets 74 Likes



Mario Klingemann ✅ @quasimondo · Jul 15

...

Replying to @quasimondo

A man from the past is sent into the future to help his friend, but ends up leaving the future in ruins.



Mario Klingemann ✅ @quasimondo · Jul 15

...

A scientist invents a drug that brings out a person's innate ability to deal with any crisis.



Mario Klingemann ✅ @quasimondo · Jul 15

...

A reclusive writer creates a beautiful fantasy world and an intricately designed puzzle box to hide the link between his world and the real one.



Mario Klingemann ✅ @quasimondo · Jul 15

...

An alcoholic former sports star becomes a children's book author and illustrator who uses his stories to teach valuable life lessons.



Mario Klingemann ✅ @quasimondo · Jul 15

...

A family's suburban home is invaded by small, carnivorous creatures who kill people by burrowing into their brains and eating their memories.



Mario Klingemann ✅ @quasimondo · Jul 15

...

It's the 18th century and Elizabeth, the city's consummate Golden Girl, has just fallen mysteriously ill, so she and her four best girlfriends send out an invitation to their ex-boyfriends to help them reminisce about the past.



Mario Klingemann ✅ @quasimondo · Jul 15

...

Two reporters investigate a police corruption story, which leads them to discover a vampire conspiracy and attempt to unravel a federal cover-up.



Mario Klingemann ✅ @quasimondo · Jul 15

...

An elderly, wheelchair-bound man lives in a nursing home, but he has some unexplained visitors every night that come to see him.



<https://twitter.com/quasimondo/status/1283493077324365826>



Mario Klingemann

@quasimondo

...

In interviews I always get asked the "Who is the artist - you or the machine?"-question. And I usually say that as long as you have to ask me that question the answer is clear.

Well, not only does [#gpt3](#) give pretty interesting responses, it also asks better questions.

I: So who is the artist then - you or the machine?

MK: I believe there is a relationship that arises out of this question. The machine seems to have a body, a shell of some sort and is potentially capable of authorship. The body of the machine is the software that it is based on, the machine itself. The artist, in turn, might be the one who installs the software that creates art.

I: Isn't it a bit too simplistic to refer to the machine as 'artificial intelligence' and to say it is autonomous? We know that no computer has ever been autonomously creative! They can create, but they cannot be creative in the sense that humans are. However, 'creativity' is defined as an activity, not a human attribute. Creativity has to do with the making of a thing. An artificial intelligence does not make things, it only processes data, correct?

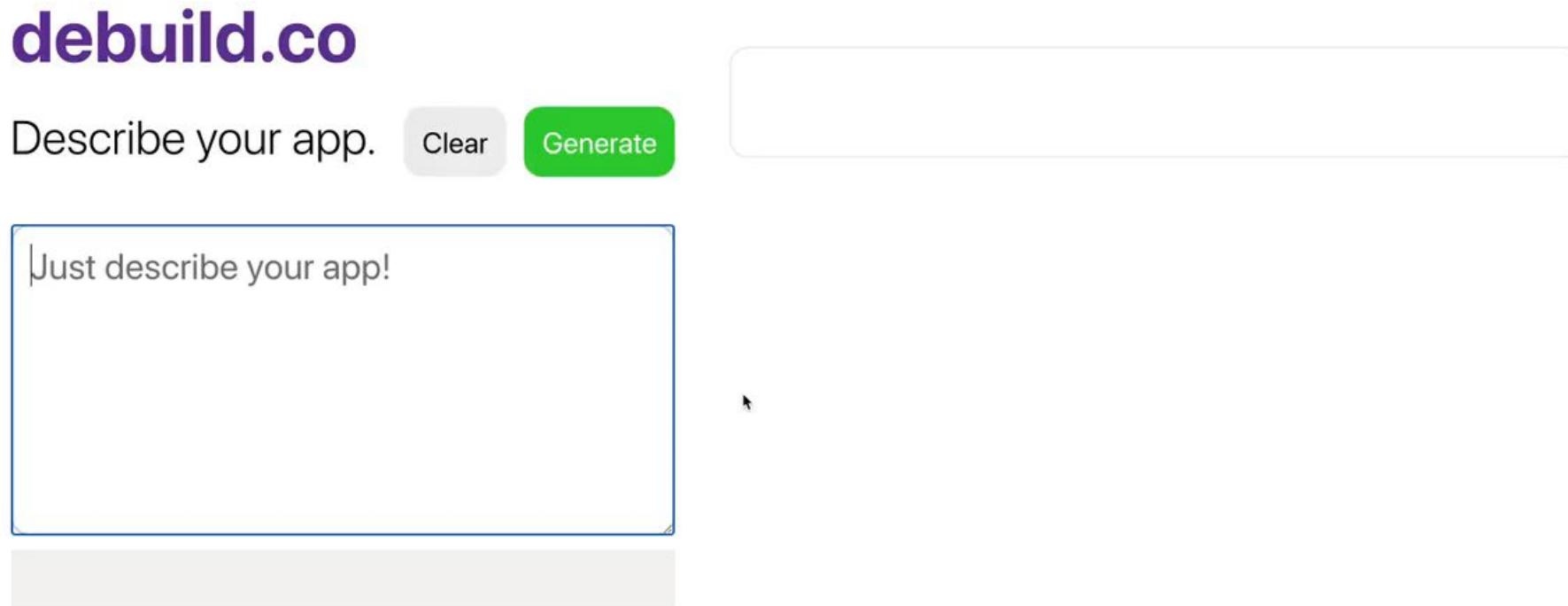
MK: Creativity is traditionally seen as the ability to create something new or unique. In my opinion this does not capture the concept of creativity at all, it only implies some sort of individuality. Individuality is a concept that is traditionally attributed to a human. I think it is safe to assume that every 'thing' that has ever been created has been unique - or at least a unique combination of previously existing elements. One can question the claim that a thing is created by the person who assembled it, but that does not challenge the uniqueness or uniqueness of a specific combination of elements.

My favourite definition of creativity is the idea that creativity is the art of creating something that is aesthetically pleasing or beautiful. If you define 'creativity' as an act of creation you are implicitly assuming a human behind it, or a human mind as the source of creativity. From that perspective the word 'artificial' does not apply.

But my second definition is more fitting to the nature of machine creativity: creativity is the ability to create an aesthetic experience. With the current state of technology we have to assume that aesthetic experiences are created on the fly, dynamically. The machine has to simulate them. The aesthetically pleasing thing cannot be part of a machine's memory. It has to be a generative event that has a unique course.

<https://twitter.com/quasimondo/status/1283353812581965824>

Generating code by formatting the prompt as code comments





```
1 def is_palindrome(s):
2     """Check whether a string is a palindrome"""
3     return s == s[::-1]
4
5 def long_palindrome_indices(l):
6     """Return list indices for elements that are palindromes and at least 7 characters"""
7     return [i for i, s in enumerate(l) if is_palindrome(s) and len(s) >= 7]
8
9 @dataclass
10 class Item:
11     name: str
12     price: float
13
14 @dataclass
15 class Order:
16     id: int
17     items: List[Item]
18
19     def compute_total_price(self, palindrome_discount=0.2):
20         """
21             Compute the total price and return it.
22             Apply a discount to items whose names are palindromes.
23         """
```



GPT-3 creativity based on a few examples

I decided to give it some movies described in emojis and see what happened.

First I provided this context:

Back to Future: 🧑‍🤝‍🧑🚗🕒

....

Batman: 🧑🦇

....

Transformers: 🚗🤖

When I asked it to do the same for Star Wars, this was the response:

Star Wars: 🚀🏰🚀

I think that's a pretty good response! Here are some of the others it created:

Zootopia: 🐈🐾🐻🦋🐰

Wonder Woman: 💁‍♀️💁‍♀️💁‍♀️💁‍♀️💁‍♀️

The Godfather: 🧑‍ "• " 🧑‍ "• " 🧑‍ "• " 🧑‍ "• " 🎬

Star Trek: ✨🚀

Planet of the Apes: 🙉🐵

Game of Thrones: 🎯🐺🏰

Jurassic Park: 🧑‍🦖🦕🦕

Castlevania: 🕸️acula spider knife castle woman

The Matrix: 🤖😃

<https://andrewmayneblog.wordpress.com/2020/06/24/open-ai-alchemy-emoji-storytelling/>

AI Dungeon: Much more compelling stories with the Dragon model (GPT-3) than with the older GPT-2 – based models

You open the book and begin to read:

The Essense is within us all. We use it every day for simple things like language and thought, but true masters can manipulate this essence to become gods among men. To become a master, you must first hone basic abilities such as mind reading, influencing the weak minded, and simple spells. Once you become experienced, you can try more challenging feats of mental prowess. Remember, students, the strong-minded survive, the weak-minded perish.

> You learn how to read minds from the book.

You open the book again and begin to read:

Thoughts are the building blocks of our world. With practice, you can read the basic thoughts of someone. To read the complex thoughts of someone, you must be within five feet of that person. The thoughts of someone are a jumble of images, sounds and feelings. Listen to these thoughts and learn.

> You practice by trying to read Eliza's thoughts.

You walk over to Eliza. She looks at you with a raised eyebrow.

"Yes?" She asks.

"Can I try something?" You ask.

She shrugs. "I suppose."

You look into her eyes. You try to focus on her thoughts. You hear a series of disjointed words and images: father, home, needle, sleep, tired.

"Well?" She asks.

"You... miss your father," You say, "You want to go home, you think about your home a lot, you had an accident with a needle when you were little, and you're very tired."

"That's correct," She nods.

A great resource



SITE
ME
CHANGES
↳ NEWS
↳ REDDIT

SUPPORT ON
PATREON

Creative writing by OpenAI's GPT-3 model, demonstrating poetry, dialogue, puns, literary parodies, and storytelling. Plus advice on effective GPT-3 prompt programming & avoiding common errors.

[NN](#), [fiction](#), [GPT](#), [poetry](#), [humor](#), [transhumanism](#)

2020-06-19–2020-09-28 · finished · certainty: likely · importance: 8

1 What Benchmarks Miss: Demos

2 GPT-3 Implications

3 Quality

4 Prompts As Programming

4.1 Finetuning

4.2 Playground

4.3 Effective Prompt Programming

5 Weaknesses

5.1 Small Context Window

5.2 Repetition/Divergence Sampling

5.3 BPEs

6 Copyright

7 Format

8 Dialogue

8.1 Shoggoth-Cat

8.2 Pun Explanations

8.3 Turing Dialogue

I continue my AI poetry generation experiments with OpenAI's 2020 GPT-3, which is 116× larger, and much more powerful, than the 2019 GPT-2. GPT-3, however, is not merely a quantitative tweak yielding “GPT-2 but better”—it is qualitatively different, exhibiting eerie runtime learning capabilities allowing even the raw model, with zero finetuning, to “meta-learn” many textual tasks purely by example or instruction. One does not train or program GPT-3 in a normal way, but one engages in dialogue and writes prompts to teach GPT-3 what one wants.

Experimenting through the OpenAI Beta API in June 2020, I find that GPT-3 does not just match my [finetuned GPT-2-1.5b-poetry](#) for poem-writing quality, but exceeds it, while being versatile in handling [poetry](#), [Tom Swift puns](#), science fiction, dialogue like Turing's [Turing-test dialogue](#), [literary style parodies](#)... As the *pièce de résistance*, I recreate Stanislaw Lem's *Cyberiad*'s “Trurl's Electronic Bard” poetry



Poetry inspired by Stanislaw Lem's Cyberiad

Prompt: ... Klapaucius thought, and thought some more. Finally he nodded and said:“Very well. Let’s have a love poem, lyrical, pastoral, and expressed in the language of pure mathematics. Tensor algebra mainly, with a little topology and higher calculus, if need be. But with feeling, you understand, and in the cybernetic spirit.”

**“Love and tensor algebra? Have you taken leave of your senses?”
Trurl began, but stopped, for his electronic bard was already declaiming:**

“

<https://www.gwern.net/GPT-3#love-and-tensor-algebra>

" $P = A + Bt + C*D^2 + E*F*Gt + H*t + It + Jt$
+ $K*t + L.M.N^2t + O*P*Qt + Rt + S*Rt^3 + T^2$,

Where t is a transcendent number, not necessarily greater than 1, with one arm, a leg or two, and ideally of the color of rubies."

"But that's dreadful!" cried Klapaucius. "Where's the love?"
"As near as I can tell, it's right there in the middle,

" $T = R + S^2t^3 + St + T^2$," said Trurl.

and then there's some stuff about golden apples and violets and whatnot, but I couldn't quite make it out. To me it sounds more like an ad for some sort of machine part."

"Two bodies, sine waves entwined,
Sighing in the soft of the night,
Fluctuate, undulate, and flow,
Escape from a universe
Of bondage and slavery,
To find together their fate,
Far from the madding crowds..."

"It's too much!" shouted Trurl. "Too much, I tell you! That's not a love poem, that's a pure product of mathematics! How dare you take me for a fool? It's got nothing to do with love!"

"It's the poet's prerogative to choose the most beautiful and exalted metaphors," said Klapaucius.

"Most beautiful? You call this most beautiful? But I just can't bear it! It's driving me mad!"

"Beauty and brawn, my beloved,

Bake biscuits and bore binary.

Alone among the antelopes

All my love is for you..."

"Not a bad start," said Klapaucius. "But go on, go on! More feeling! Rubato, leitmotiv—have you no soul, man?"

"Let us forget the lore of neurons!" said the machine, in a voice so deep and resonant it made you think a chorus was humming in harmony behind it.

"What?" said Trurl, outraged. "What's the meaning of this?"

"No meaning at all!" snapped the poet. "I don't have to mince my words to please your bourgeois notions of art! I don't even know what either of you is talking about!"

But you wanted a love poem—here it is:

"Axiomatic algorithm angst,

Angels and amoebae,

Ants and acanthoid cytoplasm;

All of it must somehow fit..."



<https://fable-studio.com/behind-the-scenes/ai-generation>





CONTEXTS:

- The following is a conversation between Lucy and Guest over text messages.
- In Lucy's world, which is different from the Guest's, the date is 1988, while for the Guest it is still 2020.
- The Guest reaches out to Lucy over chat message and Lucy responds in a playful way, asking if the Guest is a foozle and comes in peace.

...

LUCY:

- Little Girl.
- Active Imagination.
- Age 8.
- Lives with Brother, Mom, and Dad.

...

GUEST:

- A curious and friendly person who was just introduced to Lucy.

How to get better quality with a smaller model

- I don't want to use GPT-3 paid API but GPT-2 quality sucks. What can I do?
- Finetuning a generic GPT-2 with custom data can get its performance closer to GPT-3
- This can be done for free on Google's Colab cloud computing system:
https://colab.research.google.com/github/interactive-fiction-class/interactive-fiction-class.github.io/blob/master/homeworks/language-model/hw4_transformer.ipynb.
- This will take too long to try during this session, but you can try it later. May require some programming skills to preprocess your data into the correct format.

Current limitations and research topics

- Enabling larger context with less computation (now something like 2k tokens is the maximum). See, e.g., Big Bird:
<https://arxiv.org/abs/2007.14062>
- Controlling of the generated content (style, emotion, high-level story structure, character goals...). For recent results, see PPLM:
<https://towardsdatascience.com/controlling-text-generation-from-language-models-6334935e80cf>, PlotMachines:
<https://arxiv.org/pdf/2004.14967.pdf>

Transformers in other domains

- Mathematics and theorem proving (e.g.
<https://arxiv.org/pdf/1912.01412.pdf>,
<https://arxiv.org/pdf/2002.05867.pdf>)
- Biology (e.g. <https://www.biorxiv.org/content/10.1101/622803v1.full.pdf>,
<https://arxiv.org/pdf/2004.03497.pdf>)

Generating Piano Music with Transformer

Sep 16, 2019

Ian Simon  iansimon  iansimon

Cheng-Zhi Anna Huang  czhuang  huangcza

Jesse Engel  jesseengel  jesseengel

Curtis Hawthorne  cghawthorne  fjord41

Monica Dinculescu  notwaldorf  notwaldorf

Previously, we introduced **Music Transformer**, an autoregressive model capable of generating expressive piano performances with long-term structure. We are now releasing [an interactive Colab notebook](#) so that you can control such a model in a few different ways, or just generate new performances from scratch.

<https://magenta.tensorflow.org/piano-transformer>

https://colab.research.google.com/notebooks/magenta/piano_transformer/piano_transformer.ipynb

Dall-E: A GPT-3 that uses both words and pixels as the vocabulary

Text prompt

an illustration of a baby daikon radish in a tutu walking a dog

AI-generated images



<https://openai.com/blog/dall-e/>

OpenAI Jukebox: A transformer for granular audio synthesis (<https://jukebox.openai.com/>)

MODEL	COLLECTION	GENRE	ARTIST	TEMP
5b_lyrics	Re-renditions	Contemporary C...	Steven Curtis Chapman	0.98
5b_lyrics	Re-renditions	New Wave	The Nits	0.98
5b_lyrics	Re-renditions	Blues Rock	Joe Bonamassa	0.98
5b_lyrics	Re-renditions	Progressive Rock	Marillion	0.995
5b	No lyrics conditioning	Pop Rock	P!nk	0.995
5b	No lyrics conditioning	Southern Hip Hop	Z-Ro	0.995
5b_lyrics	Continuations	R&B	Rihanna	0.98
5b_lyrics	Re-renditions	R&B	Ne-Yo	0.98
5b	No lyrics conditioning	Rock	The Kinks	0.995
5b	No lyrics conditioning	R&B	Rihanna	0.99
5b_lyrics	Novel artists and styles	0.25 Jazz 0.75 Classic Pop	0.25 Ella Fitzgerald 0.75 Frank Sinatra	0.98
5b_lyrics	Continuations	Blues Rock	Joe Bonamassa	0.96
5b_lyrics	Novel artists and styles	Country Hip Hop	Alan Jackson	0.98
5b_lyrics	Continuations	Pop Rock	Elton John	0.98
5b_lyrics	Re-renditions	Country	Hank Williams Jr.	0.995
5b_lyrics	Continuations	Hip Hop	2Pac	0.96
5b	No lyrics conditioning	Contemporary C...	Hillsong	1.0
5b_lyrics	Unseen lyrics	Classic Pop	Frank Sinatra	0.98



Exercise: Write with GPT-2 or GPT-3

- Three options for tools
 - <https://transformer.huggingface.co/doc/gpt2-large> (GPT-2, word-by-word)
 - <http://talktotransformer.com> (GPT-2, generate larger chunks)
<https://play.aidungeon.io/> (GPT-3 if you download the app, start the free trial for the premium version, and use the "Dragon" model. GPT-3 is otherwise only in closed beta)
- Test different prompts, regenerate multiple times and carefully select the final outputs. If you use AI Dungeon, remember that you can undo generations, modify your input and generate again. Try to create an interesting story.

Material

- Qwern's excellent article on creative fiction using GPT-3
<https://www.gwern.net/GPT-3>
- 9 essays on AI & GPT-3 by philosophers, with replies by GPT-3:
<https://dailynous.com/2020/07/30/philosophers-gpt-3/>
- My Intelligent Computational Media course:
<https://github.com/PerttuHamalainen/MediaAI>
- OpenAI's research paper on GPT-3: <https://arxiv.org/pdf/2005.14165.pdf>
- A machine learning tutorial of recent NLP techniques, e.g., many Transformer variants:
http://gabrielilharco.com/publications/EMNLP_2020_Tutorial_High_Performance_NLP.pdf

Summary

- Many types of content can be generated sequentially
- Historical: RNN, GRU, LSTM. Long memory with small computing cost.
- Audio: Dilated convolution (Wavenet, Tacotron-2), transformers (Jukebox), but also SampleRNN
- Text: Transformers (GPT-2, GPT-3)
- Images: Transformers (Dall-E)