ML Applications

Mauro Verzetti, Google AI

>>> whoami

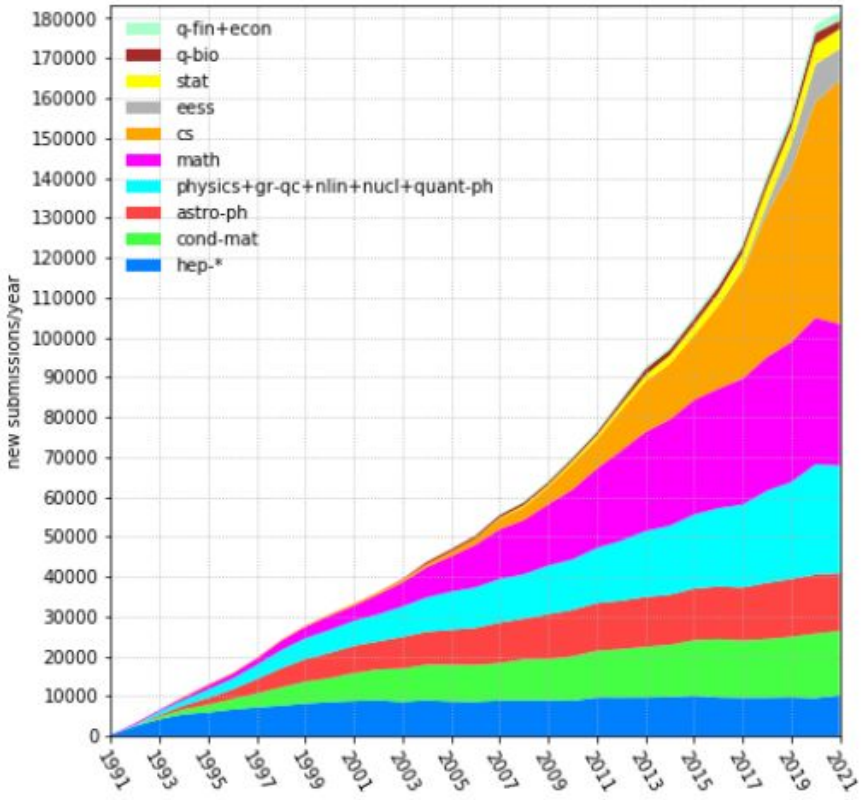
I am a HEP physicist ~recently transitioned to Google AI research

- Master thesis on Bottomonium decays @ Belle
- PhD @ UZH on WH -> tau tau @ CMS
- Post-doc @ U. Rochester
- CERN Fellow / Staff
  - Jet tagging
  - Top physics
  - B-physics (B-parking effort)
- Google AI focusing on audio
  - Ambient and music

Disclaimer #1

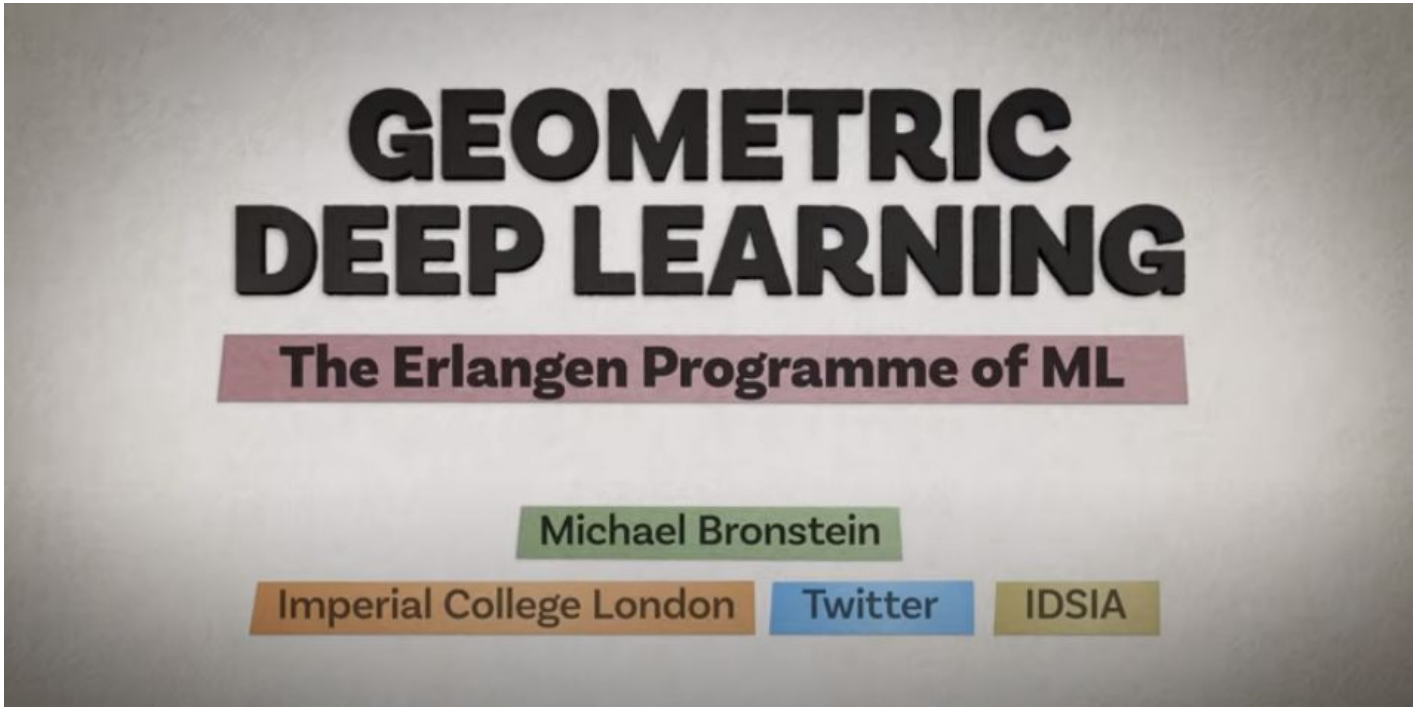
The ML field is **HUGE** and moving fast

- It is hard (if not impossible) to have a detailed view of all the developments
- My views (and this presentation) are necessarily biased by my research field and lab



From arXiv stats

A case in point



ICLR 2021 Keynote - "Geometric Deep Learning: The Erlangen Programme of ML" - M Bronstein

Disclaimer #2

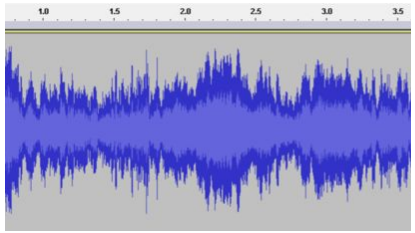
HEP

- Large **labelled** synthetic datasets
- Well established metrics
- High-precision simulations
  - Generally high accuracy is required
- High-level signals
- Low-ish compute power

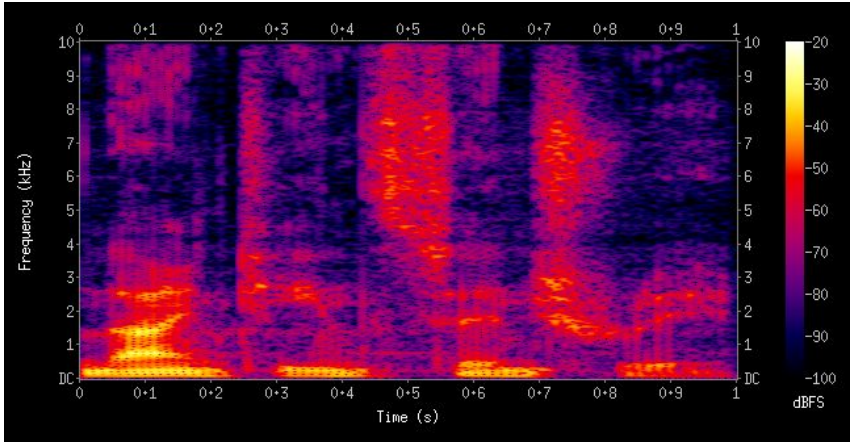
AI/ML in an industry research lab

- *Some* large labelled dataset, but generally small and noisy
- **Huge unlabelled** datasets
- Mapping human behaviours/preferences, metrics are a guidance, but not necessarily a target
- No synthetic data
- Low-level signals (pixels, waveforms/spectrograms, clicks, tokens, etc..)
- High compute power

Audio signal processing 101 - *accelerated* and condensed



$$log(|STFT(x_t, \theta)|^2)$$



Vocoder (ML)

Images from wikimedia [1], [2]

Generative models

CNBC

Search quotes, news & videos | WATCHLIST | SIG

MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV INVESTING CLUB PRO MAKE IT SELECT

TECH

All you need to know about ChatGPT, the A.I. chatbot that's got the world talking and tech giants clashing

PUBLISHED WED, FEB 8 2023•7:37 AM EST | UPDATED WED, FEB 8 2023•10:52 AM EST

Ryan Browne
@RYAN_BROWNE_

SHARE

GOOGLE / TECH / ARTIFICIAL INTELLIGENCE

Google's new AI turns text into music

Song of the bots. Illustration: The Verge

/ The examples the company has shared are music to my ears.

By MITCHELL CLARK
Jan 28, 2023, 4:00 PM GMT+1 | 13 Comments / 13 New

The Guardian

Sport Culture Lifestyle More

crisis Environment Science Global development Football Tech Business Obituaries

This article is more than 5 months old

Dall-E 2 users to be allowed to upload faces for first time

Feature marks latest relaxation of rules around how image-generating AI tool can be used

Art created by artificial intelligence: "Frightening and fascinating all at the same time."

CBS News · 15 Jan 2023

CNBC

Search quotes, news & videos | WATCHLIST | SIG

MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV INVESTING CLUB PRO MAKE IT SELECT

TECH

All you need to know about ChatGPT, the A.I. chatbot that's got the world talking and tech giants clashing

PUBLISHED WED, FEB 8 2023•7:37 AM EST | UPDATED WED, FEB 8 2023•10:52 AM EST

Ryan Browne
@RYAN_BROWNE_

SHARE

TRANSFORMERS

GOOGLE / TECH / ARTIFICIAL INTELLIGENCE

Google's new AI turns text into music

/ The examples the company has shared are music to my ears.

By MITCHELL CLARK
Jan 28, 2023, 4:00 PM GMT+1 | 13 Comments / 13 New

Song of the bots. Illustration: The Verge

The Guardian

Sport Culture Lifestyle More

crisis Environment Science Global development Football Tech Business Obituaries

This article is more than 5 months old

Dall-E 2 users to be allowed to upload faces for first time

Feature marks latest relaxation of rules around how image-generating AI tool can be used

DIFFUSION MODELS

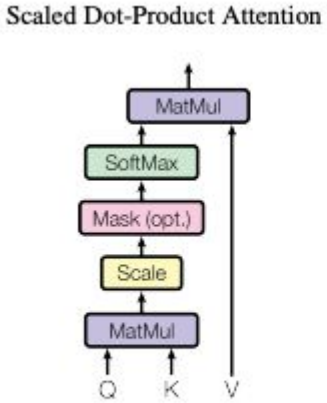
Art created by artificial intelligence: "Frightening and fascinating all at the same time."

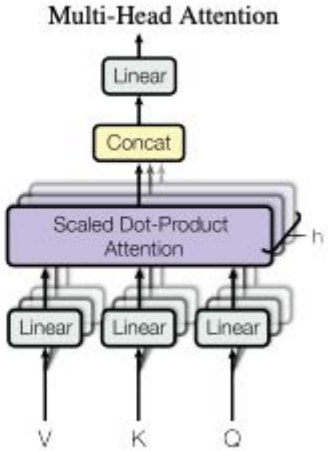
CBS News · 15 Jan 2023

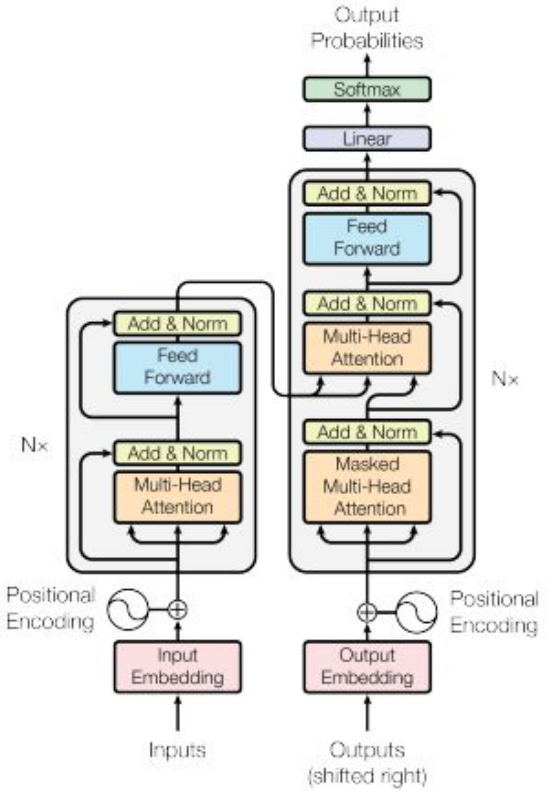
Transformers

Transformers - Attention is all you need - arxiv:1706.03762

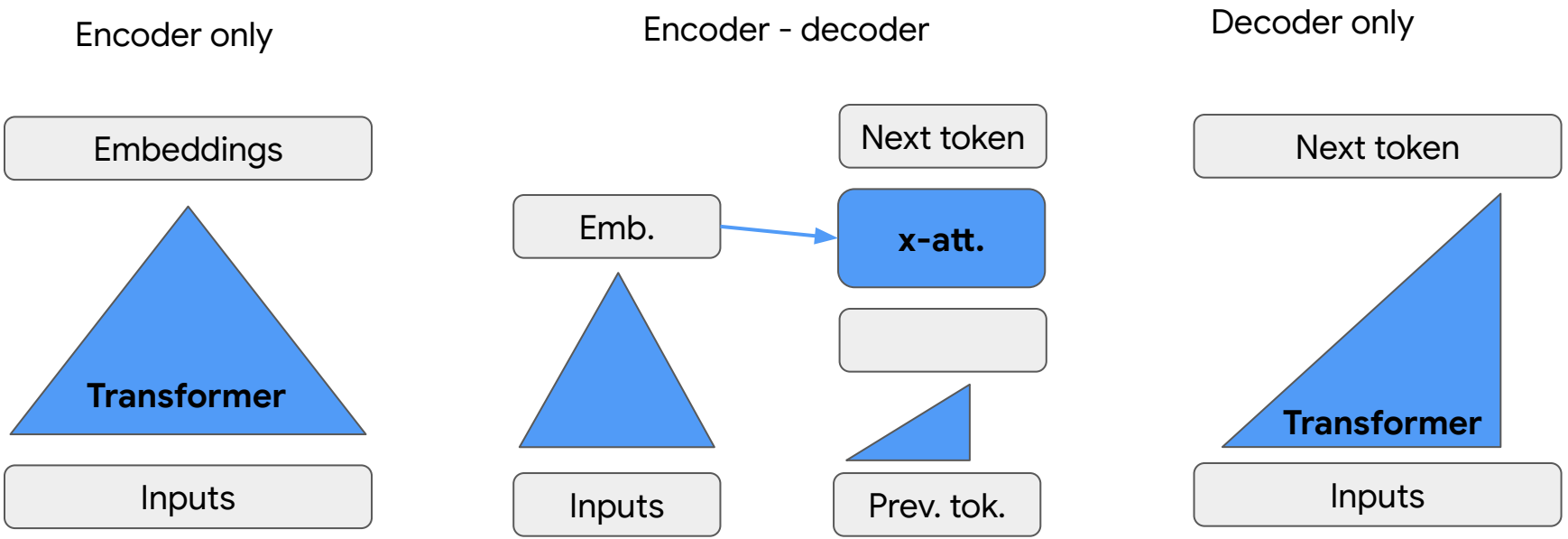
$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$





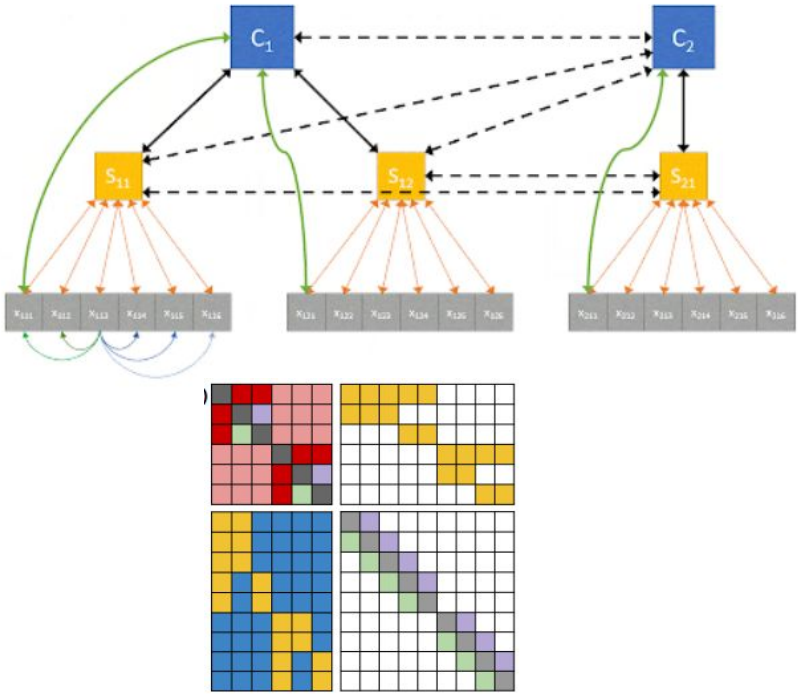


Transformers types

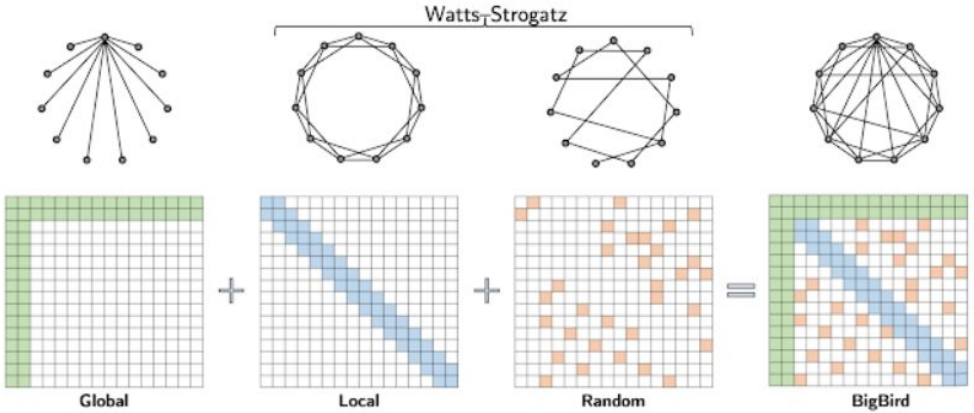


Long sequences (1) - Local attention: ETC, BigBird

arxiv:2004.08483

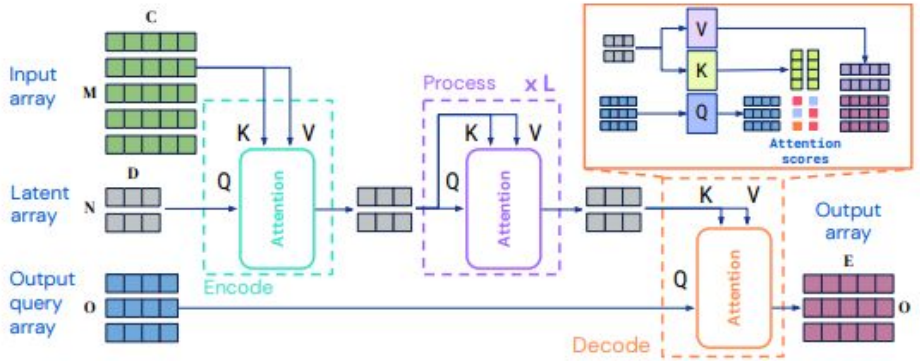


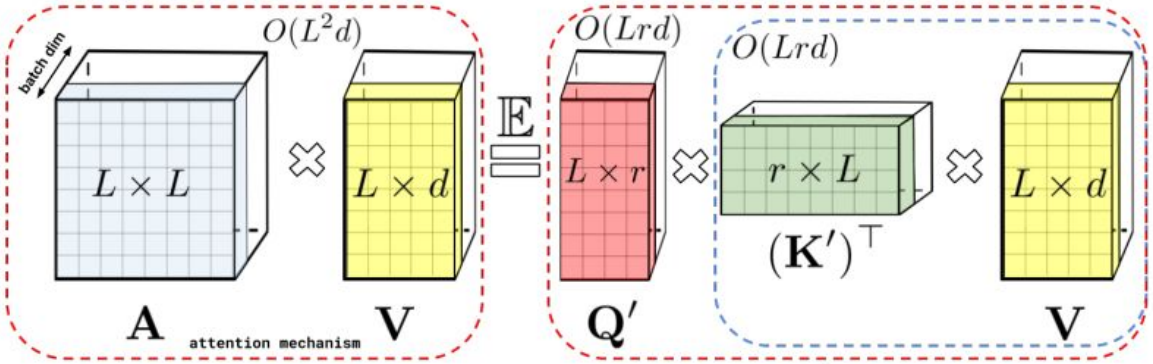
arxiv:2007.14062



Long sequences (2) - Perceiver I/O, Peformer

arxiv:2107.14795

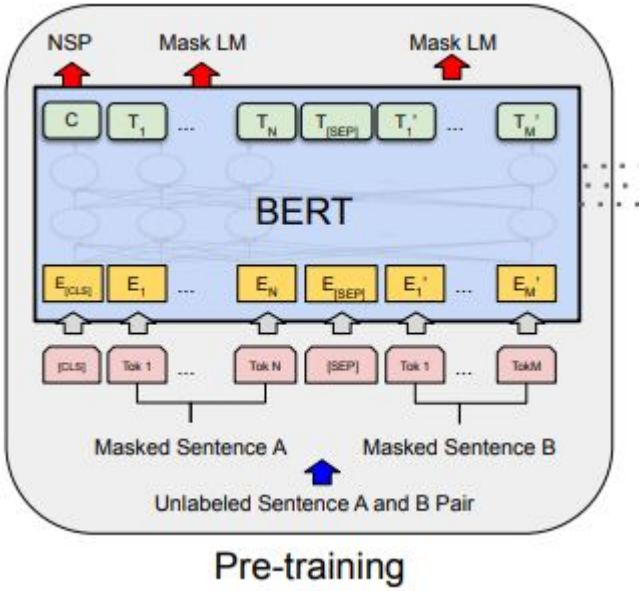




arxiv:2009.14794

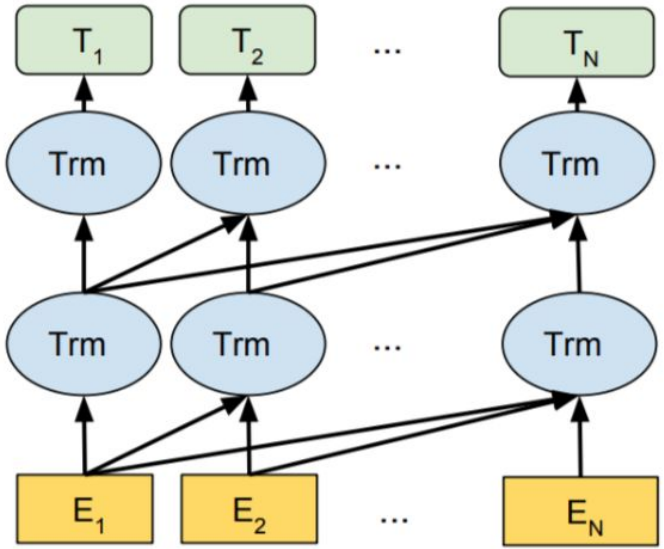
LMs, why so successful? - Training on unsupervised data

arxiv:1810.04805



Pre-training

GPT-2 [paper] [graphic]

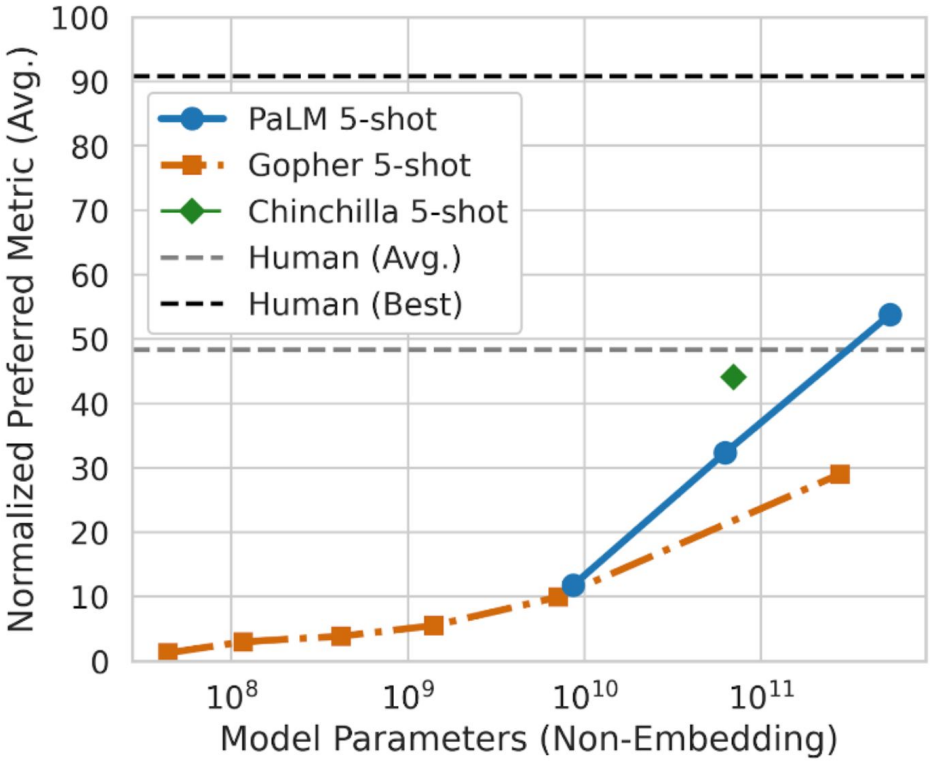


LMs got bigger: LLMs

PaLM arxiv:2204.02311
GPT-3 arxiv:2005.14165
Chinchilla arxiv:2203.15556
LaMDA arxiv:2201.08239

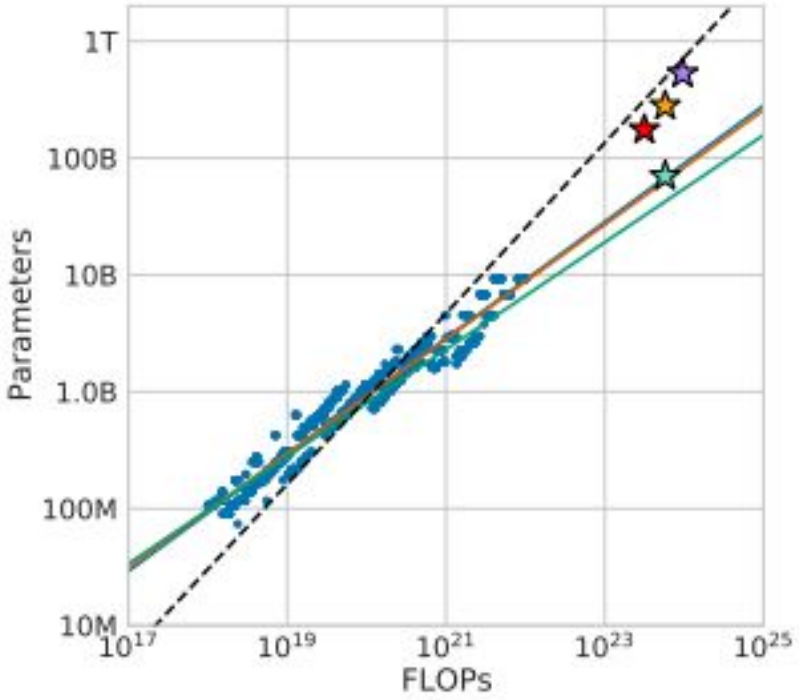
Emerging abilities

- Zero-shot learning
- Summarization
- Translation
- Mathematical abilities
- Chain of tough
- ...



"We trained PaLM-540B on 6144 TPU v4 chips for 1200 hours..."

LLMs scaling rules

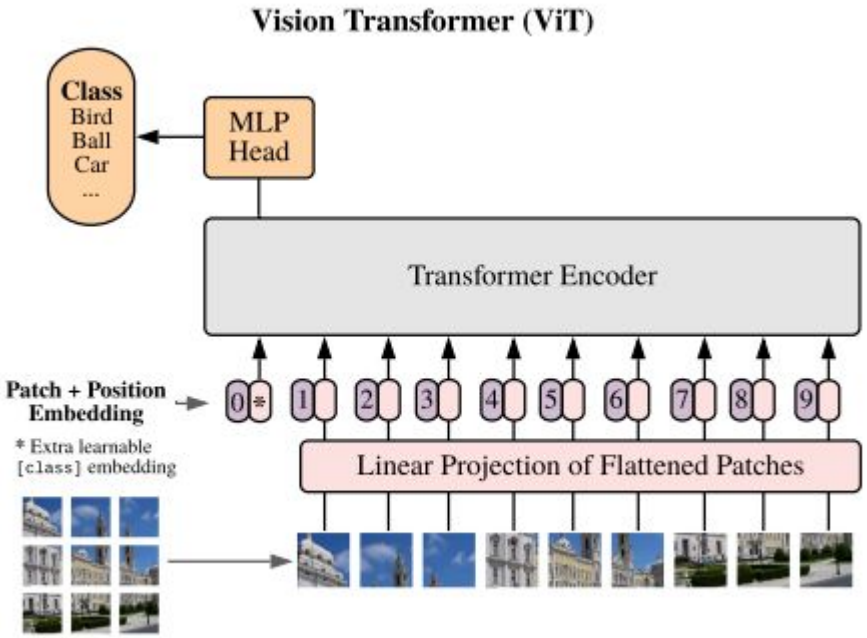


arxiv:2203.15556

Approach 1
Approach 2
Approach 3
Kaplan et al (2020)

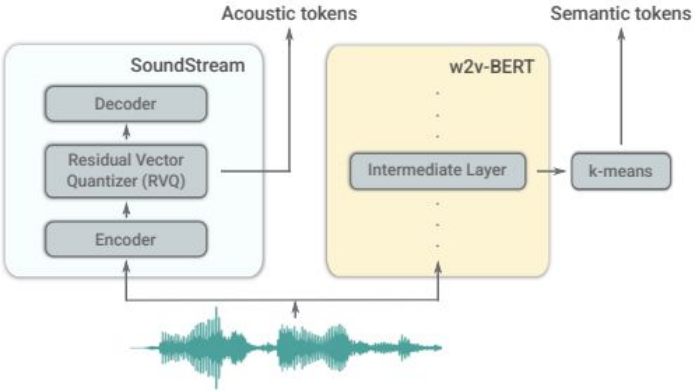
Chinchilla (70B)
Gopher (280B)
GPT-3 (175B)
Megatron-Turing NLG (530B)

Transformers outside languages: Vision Transformers



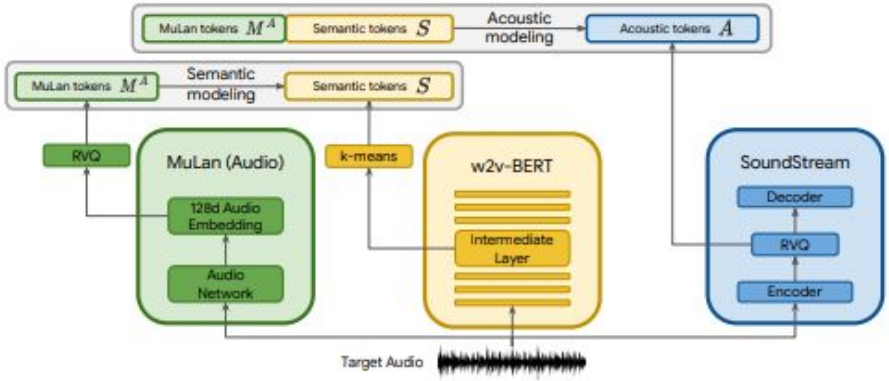
arxiv:2010.11929

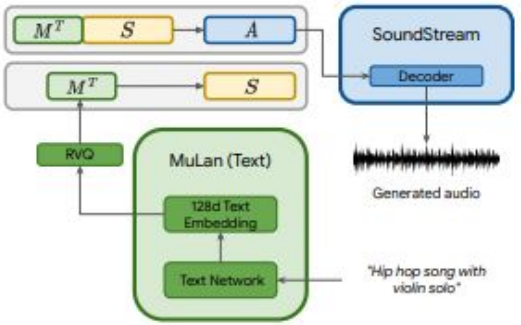
Transformers outside languages: AudioLM / MusicLM



arxiv:2209.03143
google-research.github.io/seanet/audiolm/examples/

arxiv:2301.11325
google-research.github.io/seanet/musiclm/examples/

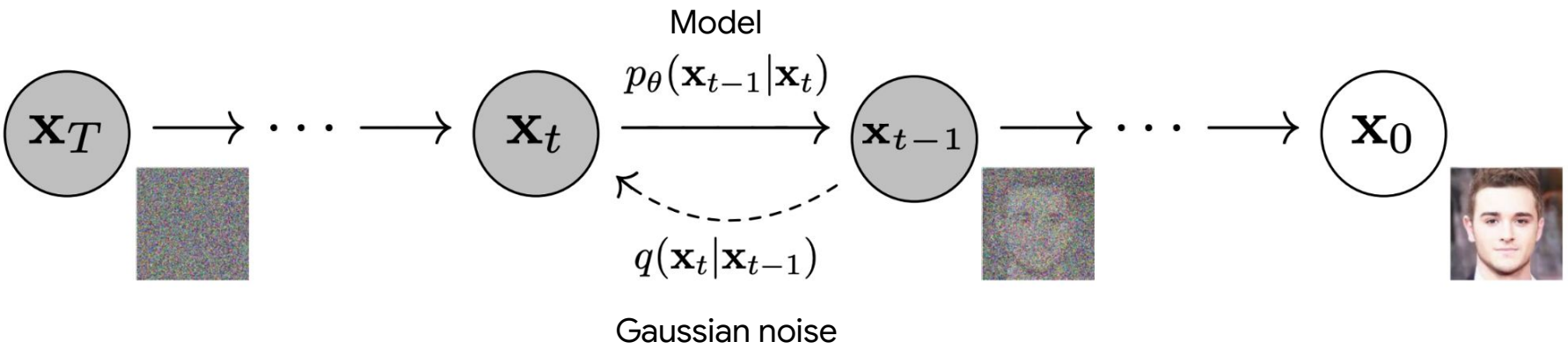




Diffusion models

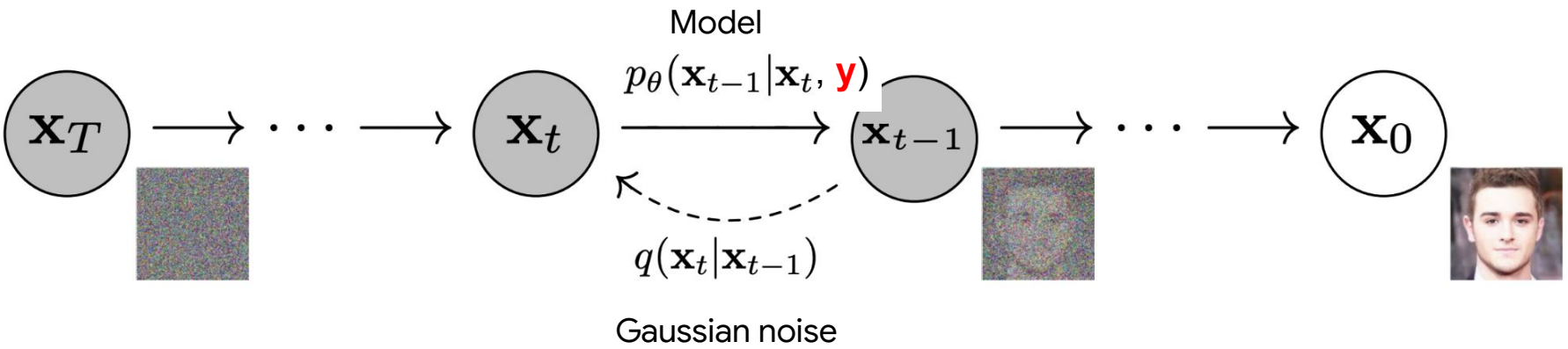
Diffusion models basics

arxiv:2006.11239

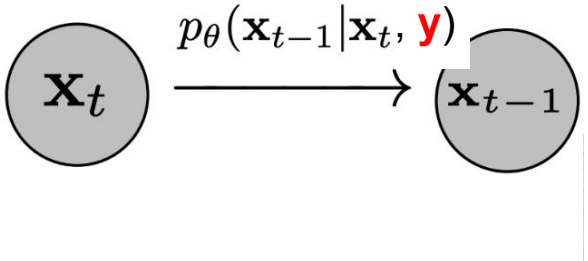


Diffusion models basics

arxiv:2006.11239



Diffusion models basics

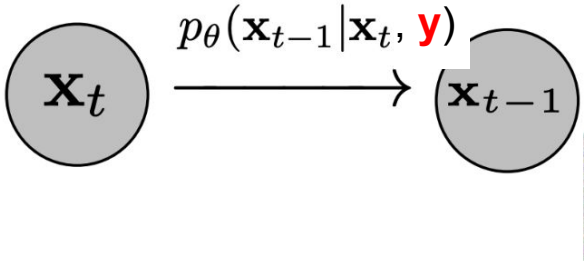


$$p_\theta(x_{t-1}|x_t, y) \approx \nabla_x p_\theta(x|y)$$

Classifier: $p_\phi(y|x) \rightarrow \nabla_x p_\phi(y|x)$

Classifier guidance: $\nabla_x p_{\theta,\phi,\gamma}(x|y) = \nabla_x p_\theta(x|y) + \gamma \nabla_x p_\phi(x|y)$ arxiv:2105.05233

Diffusion models basics



$$p_\theta(x_{t-1}|x_t, y) \approx \nabla_x p_\theta(x|y)$$

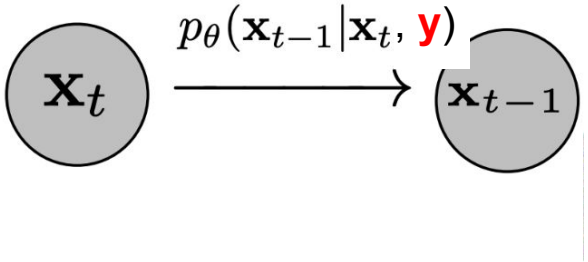
Classifier: $p_\phi(y|x) \rightarrow \nabla_x p_\phi(y|x)$

Classifier guidance: $\nabla_x p_{\theta,\phi,\gamma}(x|y) = \nabla_x p_\theta(x|y) + \gamma \nabla_x p_\phi(x|y)$ arxiv:2105.05233

Classifier-***free*** guidance: $\nabla_x p_{\theta,\phi,\gamma}(x|y) = \nabla_x p_\theta(x|y) - \gamma \nabla_x p_\phi(x|-)$

arxiv:2207.12598

Diffusion models basics



$$p_\theta(x_{t-1}|x_t, y) \approx \nabla_x p_\theta(x|y)$$

Classifier: $p_\phi(y|x) \rightarrow \nabla_x p_\phi(y|x)$

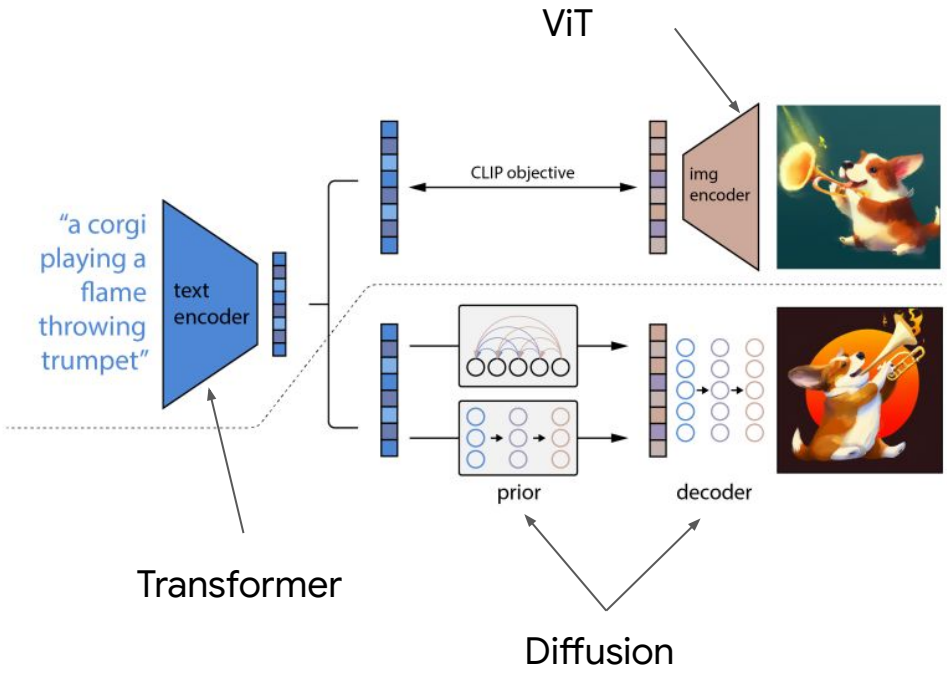
Classifier guidance: $\nabla_x p_{\theta,\phi,\gamma}(x|y) = \nabla_x p_\theta(x|y) + \gamma \nabla_x p_\phi(x|y)$ arxiv:2105.05233

Classifier-***free*** guidance: $\nabla_x p_{\theta,\phi,\gamma}(x|y) = \nabla_x p_\theta(x|y) - \gamma \nabla_x p_\phi(x|-)$

arxiv:2207.12598

Negative prompt: $\nabla_x p_{\theta,\phi,\gamma}(x|y) = \nabla_x p_\theta(x|y) - \gamma \nabla_x p_\phi(x|\mathbf{y}')$

Dall-e 2

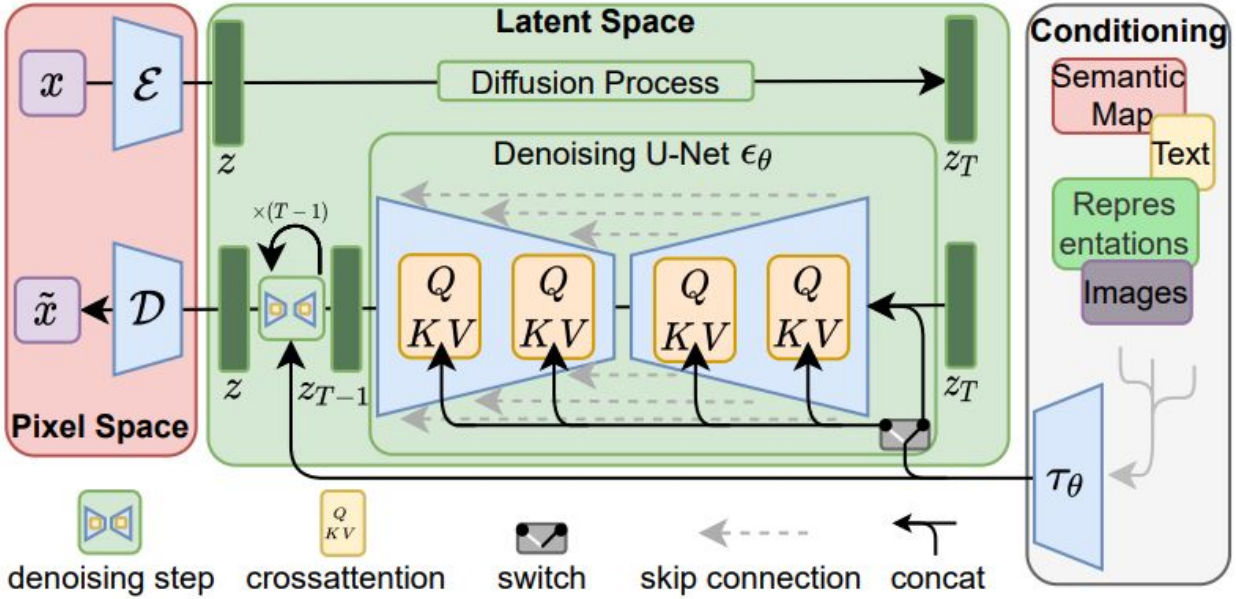


arxiv:2204.06125

vibrant portrait painting of Salvador Dalí with a robotic half face

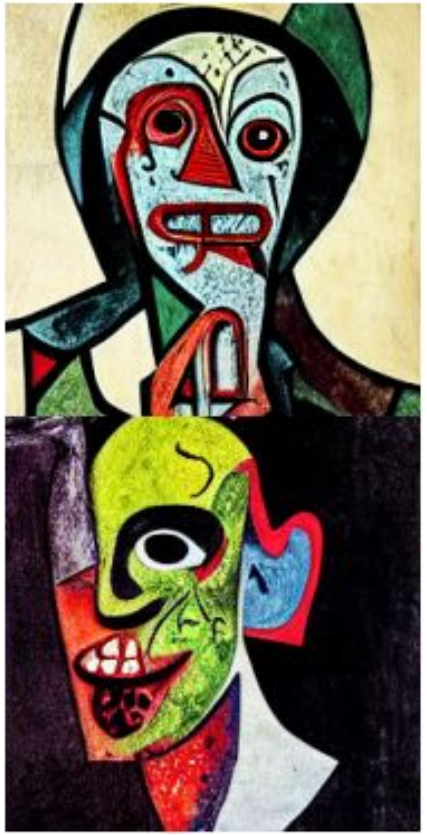
an espresso machine that makes coffee from human souls, artstation

Stable diffusion



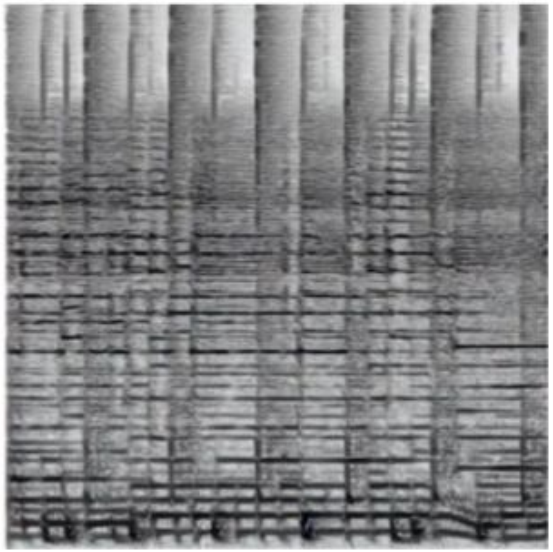
arxiv:2112.10752

'A zombie in the style of Picasso'

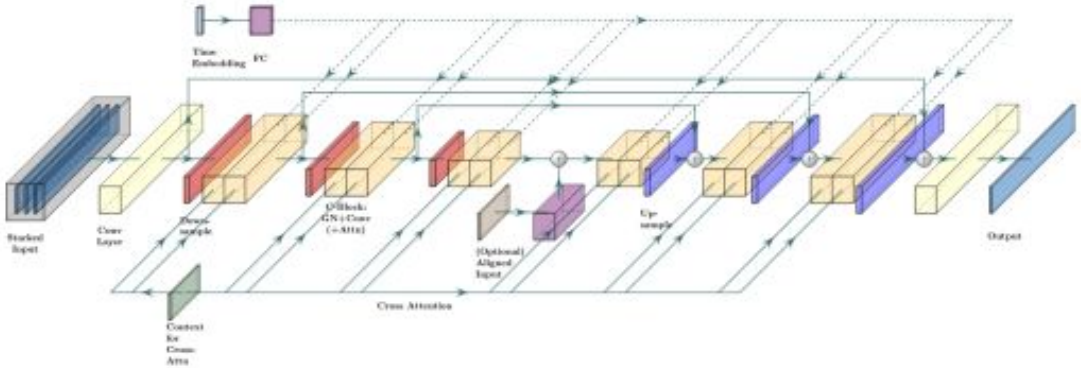


Music diffusion - Riffusion and Noise2Music

rock and roll electric guitar solo



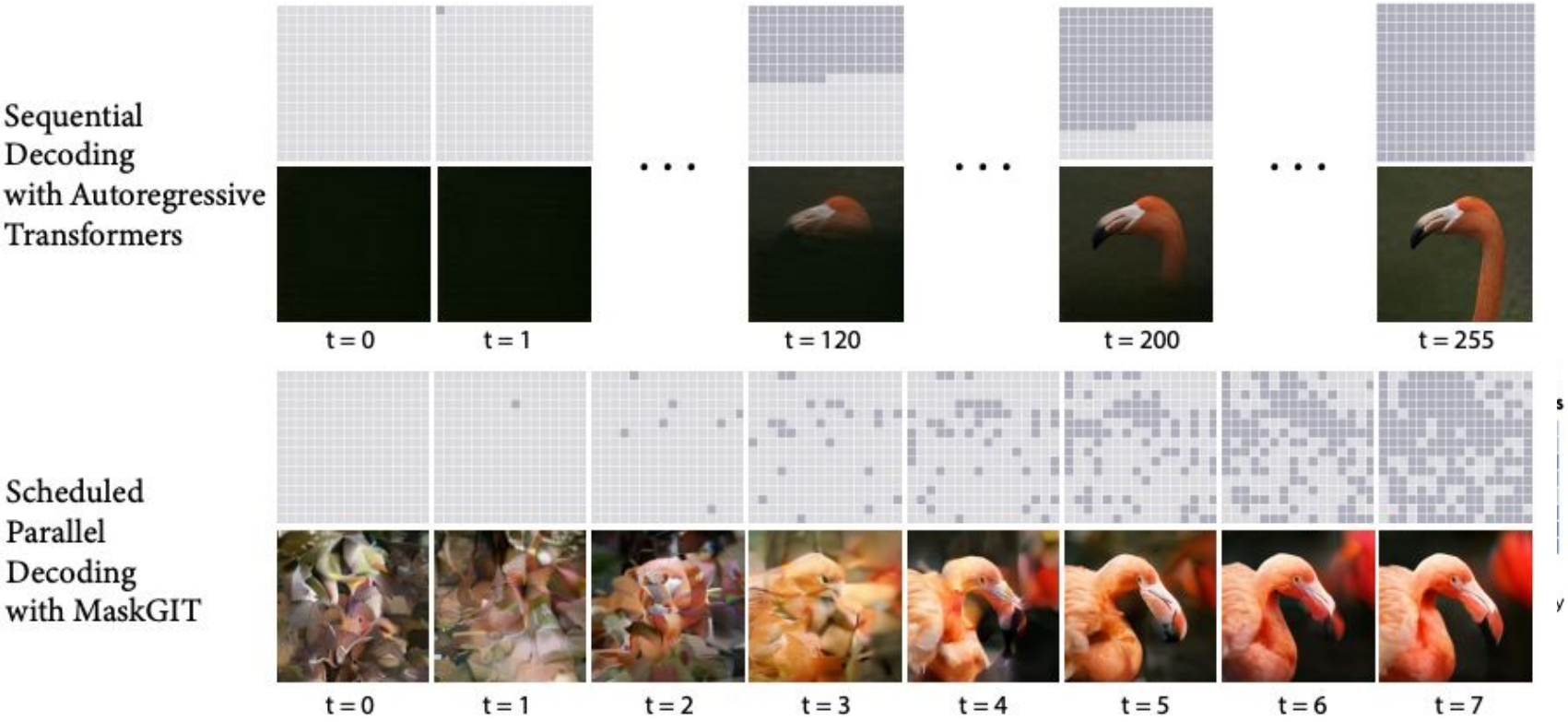
0:00 / 0:05



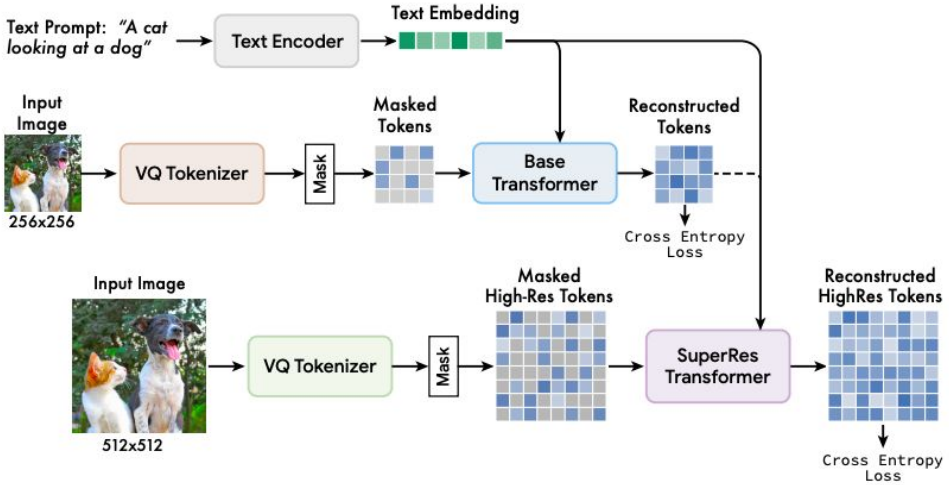
arxiv:2302.03917

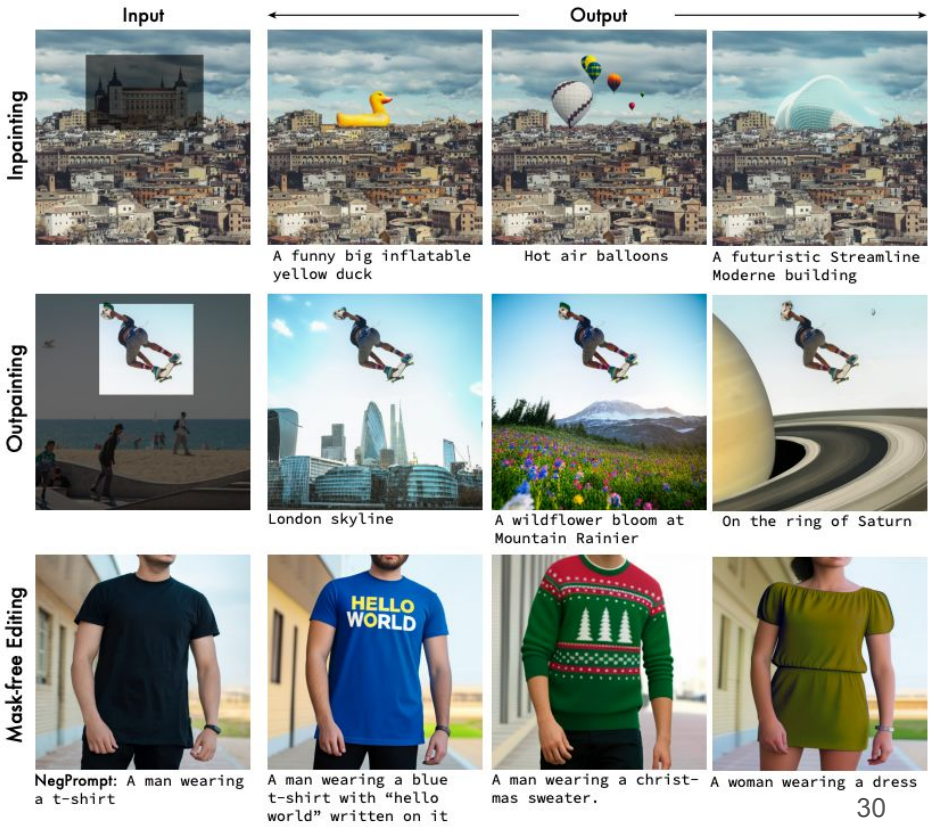
A hybrid approach - MaskGit/Muse

arxiv:2301.00704



A hybrid approach - MaskGit/Muse





arxiv:2301.00704

Conclusions

How would you conclude a presentation on the latest generative AI techniques and architectures?

In conclusion, we have seen that generative AI techniques and architectures have come a long way in recent years, from simple models such as Variational Autoencoders to complex models such as Generative Adversarial Networks and Transformers. These models have enabled us to generate high-quality images, text, and even music, which was previously impossible using traditional methods.

However, there are still several challenges that need to be addressed, such as mode collapse, training instability, and the generation of coherent long-form text. Despite these challenges, the potential applications of generative AI are vast and exciting, ranging from generating synthetic data for training models to creating new art forms and even helping in drug discovery.

As the field of generative AI continues to evolve, it is crucial to keep pushing the boundaries of what is possible while also addressing the ethical and societal implications of these technologies. With continued research and development, we can expect even more impressive and impactful applications of generative AI in the future.