# TRANSFORMER MODELS: AN INTRODUCTION AND CATALOG

Xavier Amatriain Los Gatos, CA 95032 xavier@amatriain.net

February 16, 2023

#### ABSTRACT

I have a terrible memory for names. In the past few years we have seen the meteoric appearance of dozens of models of the Transformer family, all of which have funny, but not self-explanatory, names. The goal of this post is to offer a short and simple catalog and classification of the most popular Transformer models. In other words, I needed a Transformers cheat-sheet and could not find a good enough one online, so I thought I'd write my own. I hope it can be useful to you too.

# Contents

1	Intr	oductio	n: What are Transformers	3						
	1.1	1 Encoder/Decoder architecture								
	1.2	Attent	ion	5						
	1.3	What a	are Transformers used for and why are they so popular	5						
	1.4	RLHF		6						
	1.5	Diffus	ion	6						
2	The	Transfo	ormers catalog	8						
	2.1	Featur	es of a Transformer	8						
		2.1.1	Pretraining Architecture	8						
		2.1.2	Pretraining Task	8						
		2.1.3	Application	9						
	2.2	Catalo	g table	9						
	2.3	Family	/ Tree	9						
	2.4	Chrone	ological timeline	9						
	2.5	Catalo	g List	9						
		2.5.1	ALBERT	9						
		2.5.2	AlphaFold	9						
		2.5.3	Anthropic Assistant	12						
		2.5.4	BART	13						
		2.5.5	BERT	13						

Big Bird	13
BlenderBot3	14
BLOOM	14
ChatGPT	14
Chinchilla	15
CLIP	15
CM3	15
CTRL	16
DALL-E	16
DALL-E 2	16
Decision Transformers	17
DialoGPT	17
DistilBERT	17
DQ-BART	18
ELECTRA	18
ERNIE	18
Flamingo	19
Gato	19
GLaM	19
GLIDE	20
Global Context ViT	20
Gopher	20
GopherCite	21
GPT	21
GPT-2	21
GPT-3	22
GPT-3.5	22
InstructGPT	22
GPT-Neo	23
GPT-NeoX-20B	23
HTML	23
Imagen	23
Jurassic-1	24
LAMDA	24
mBART	24
Megatron	25
Minerva	25
MT-NLG (Megatron TouringNLG)	25
OPT	26
	Big But   BlenderBot3   BLOOM   ChatGPT   Chinchilla   CLIP   CM3   CTRL   DALL-E   DALL-E 2   Decision Transformers   DialoGPT   DistilBERT   DQ-BART   ELECTRA   ERNIE   Flamingo   Gato   Global Context VIT   GopherCite   GPT-3   GPT-8   GPT-NeoX-20B   HTML   Imagen   Jurassic-1   LAMDA   Minerva   Mireva   Minerva

2.5.45	PalM	26
2.5.46	Pegasus	26
2.5.47	RoBERTa	27
2.5.48	SeeKer	27
2.5.49	Sparrow	27
2.5.50	StableDiffusion	28
2.5.51	Swin Transformer	28
2.5.52	Switch	28
2.5.53	T5	29
2.5.54	Trajectory Transformers	29
2.5.55	Transformer XL	29
2.5.56	Turing-NLG	30
2.5.57	ViT	30
2.5.58	Wu Dao 2.0	30
2.5.59	XLM-RoBERTa	31
2.5.60	XLNet	31
her rea	ding	31

# 3 Further reading

# **1** Introduction: What are Transformers

Transformers are a class of deep learning models that are defined by some architectural traits. They were first introduced in the now famous "Attention is All you Need" paper by Google researchers in 2017 [1] (the paper has accumulated a whooping 38k citations in only 5 years) and associated blog post<sup>1</sup>.

The Transformer architecture is a specific instance of the encoder-decoder models  $[2]^2$  that had become popular just over the 2–3 years prior. Up until that point however, attention was just one of the mechanisms used by these models, which were mostly based on LSTM (Long Short Term Memory)[3] and other RNN (Recurrent Neural Networks)[4] variations. The key insight of the Transformers paper was that, as the title implies, attention could be used as the only mechanism to derive dependencies between input and output.

It is beyond the scope of this blog to go into all the details of the Transformer architecture. For that, I will refer you to the original paper above or to the wonderful The Illustrated Transformer<sup>3</sup> post. That being said, we will briefly describe the most important aspects since we will be referring to them in the catalog below. Let's start with the basic architectural diagram from the original paper, and describe some of the components.

# **1.1 Encoder/Decoder architecture**

A generic encoder/decoder architecture (see Figure 1) is made up of two models. The encoder takes the input and encodes it into a fixed-length vector. The decoder takes that vector and decodes it into the output sequence. The encoder and decoder are jointly trained to minimize the conditional log-likelihood. Once trained the encoder/decoder can generate an output given an input sequence or can score a pair of input/output sequences.

In the case of the original Transformer architecture, both encoder and decoder had 6 identical layers. In each of those 6 layers the Encoder has two sub layers: a multi-head attention layer, and a simple feed forward network. Each sublayer has a residual connection and a layer normalization. The output size of the Encoder is 512. The Decoder adds a third sublayer, which is another multi-head attention layer over the output of the Encoder. Besides, the other multi-head layer in the decoder is masked to prevent attention to subsequent positions.

<sup>&</sup>lt;sup>1</sup>https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html

<sup>&</sup>lt;sup>2</sup>https://machinelearningmastery.com/encoder-decoder-long-short-term-memory-networks/

<sup>&</sup>lt;sup>3</sup>https://jalammar.github.io/illustrated-transformer/



Figure 1: Transformer Architecture from [1]



Figure 2: (left) Scaled Dot-Product Attention. (right) Multi-Head Attention consists of several attention layers running in parallel.

Figure 2: The Attention Mechanism

#### 1.2 Attention

It is clear from the description above that the only "exotic" elements of the model architecture are the multi-headed attention, but, as described above, that is where the whole power of the model lies! So, what is attention anyway? An attention function is a mapping between a query and a set of key-value pairs to an output. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key. Transformers use multi-headed attention, which is a parallel computation of a specific attention function called scaled dot-product attention. I will refer you again to the The Illustrated Transformer<sup>4</sup> post for many more details on how the attention mechanism works, but will reproduce the diagram from the original paper in Figure 2 so you get the main idea

There are several advantages of attention layers over recurrent and convolutional networks, the two most important being their lower computational complexity and their higher connectivity, especially useful for learning long-term dependencies in sequences.

#### 1.3 What are Transformers used for and why are they so popular

The original transformer was designed for language translation, particularly from English to German. But, already the original paper showed that the architecture generalized well to other language tasks. This particular trend became quickly noticed by the research community. Over the next few months most of the leaderboards for any language-related ML task became completely dominated by some version of the transformer architecture (see for example the well known SQUAD leaderboard<sup>5</sup> for question answer where all models at the top are ensembles of Transformers).

One of the key reasons Transformers were able to so quickly take over most NLP leaderboards is their ability to quickly adapt to other tasks, a.k.a. Transfer learning. Pretrained Transformer models can adapt extremely easily and quickly to tasks they have not been trained on, and that has huge advantages. As an ML practitioner, you no longer need to train a large model on a huge dataset. All you need to do is re-use the pretrained model on your task, maybe just slightly

<sup>&</sup>lt;sup>4</sup>https://jalammar.github.io/illustrated-transformer/

<sup>&</sup>lt;sup>5</sup>https://rajpurkar.github.io/SQuAD-explorer/

adapting it with a much smaller data set. A specific technique used to adapt pretrained models to a different task is the so-called fine tuning<sup>6</sup>.

It turns out that the capability of Transformers to adapt to other tasks is so great, that, while they were initially developed for language related tasks, they quickly became useful for other tasks ranging from vision[5] or audio and music<sup>7</sup> applications all the way to playing chess[6] or doing math[7].

Of course all these applications would have not been possible if it wasn't because of the myriad of tools that made them readily available to anyone that could write a few lines of code. Not only were Transformers quickly integrated into the main AI frameworks (namely Pytorch<sup>8</sup> and TF<sup>9</sup>), but they even enabled the creation of an entire company around them. Huggingface<sup>10</sup>, a startup that has raised over \$ 60M to this day, is almost entirely built around the idea of commercializing their open source Transformers library<sup>11</sup>.

Last but not least, I would be remiss if I did not mention the impact of GPT-3[8] on the popularization of Transformers. GPT-3 is a Transformer model introduced by OpenAI in May 2020 as a follow up to their earlier GPT and GPT-2. The company made a big splash by introducing the model in a preprint[8] in which they claimed that the model was so powerful that they were not in a position to release it to the world. Since then, the model has not only been released, but also commercialized through a very large partnership<sup>12</sup> between OpenAI and Microsoft. GPT-3 powers over 300 different applications<sup>13</sup>, and is the foundation for OpenAI's commercial strategy (which is a lot to say for a company that has received over \$ 1B in funding).

# 1.4 RLHF

Reinforcement Learning from Human Feedback (or Preferences) aka RLHF (or RLHP) has become a huge addition to the AI toolkit as of lately. The concept was introduced already in 2017 in the paper ["Deep reinforcement learning from human preferences"](https://arxiv.org/abs/1706.03741). More recently though, it has been applied to ChatGPT and similar dialog agents like BlenderBot3 or Sparrow. The idea is pretty simple though: Once a language model is pretrained, we can generate different responses to a dialog and have Humans rank the results. We can use those ranking (aka preferences or feedback) to train a reward, in the reinforcement learning context (see Figure 3). You can read much more in these two wonderful posts by Huggingface]<sup>14</sup> or Weights and Bias<sup>15</sup>.

# 1.5 Diffusion

Diffusion models have become the new SOTA in image generation, clearly pushing aside the previous approaches such as GANs (Generative Adversarial Networks). What are diffusion models? They are a class of latent variable models trained variational inference. What this means in practice is that we train a deep neural network to denoise images blurred with some sort of noise function. Networks that are trained this way are in fact learning the latent space of what those images represent (see figure 4.

Diffusion models have relation with other generative models like the famous [Generative Adversarial Networks (GAN)]<sup>16</sup>, which they have mostly replaced in many applications and, particularly with (denoising) Autoencoders. Some authors<sup>17</sup> will go as far as saying that Diffusion models are just a specific instance of autoencoders. However, they also admit that the small differences do transform their application, from the latent representation of autoconders to the pure generative nature of Diffusion models.

<sup>&</sup>lt;sup>6</sup>https://huggingface.co/docs/transformers/training

<sup>&</sup>lt;sup>7</sup>https://magenta.tensorflow.org/music-transformer

<sup>&</sup>lt;sup>8</sup>https://pytorch.org/tutorials/beginner/transformer\_tutorial.html

<sup>&</sup>lt;sup>9</sup>https://www.tensorflow.org/text/tutorials/transformer

<sup>&</sup>lt;sup>10</sup>https://huggingface.co/docs

<sup>&</sup>lt;sup>11</sup>https://github.com/huggingface/transformers

<sup>&</sup>lt;sup>12</sup>https://openai.com/blog/openai-licenses-gpt-3-technology-to-microsoft/

<sup>&</sup>lt;sup>13</sup>https://openai.com/blog/gpt-3-apps/

<sup>&</sup>lt;sup>14</sup>https://huggingface.co/blog/rlhf

<sup>&</sup>lt;sup>15</sup>https://wandb.ai/ayush-thakur/RLHF/reports/Understanding-Reinforcement-Learning-from-Human-Feedback-RLHF-Part-1

<sup>&</sup>lt;sup>16</sup>https://en.wikipedia.org/wiki/Generative\_adversarial\_network

<sup>&</sup>lt;sup>17</sup>https://benanne.github.io/2022/01/31/diffusion.html



Figure 3: Reinforcement Learning with Human Feedback. From HuggingFace's RLHF blog post at https://huggingface.co/blog/rlhf.



Fig. 2. Diffusion models smoothly perturb data by adding noise, then reverse this process to generate new data from noise. Each denoising step in the reverse process typically requires estimating the score function (see the illustrative figure on the right), which is a gradient pointing to the directions of data with higher likelihood and less noise.

Figure 4: Probabilistic diffusion model architecture from "Diffusion Models: A Comprehensive Survey of Methods and Applications" [9]

# 2 The Transformers catalog

**Note:** For all the models available in Huggingface, I decided to directly link to the page in the documentation since they do a fantastic job of offering a consistent format and links to everything else you might need, including the original papers. Only a few of the models are not included in Huggingface. For those, I try to include a link to their github if available or blog post if not. For all, I also include bibliographic reference.

# 2.1 Features of a Transformer

So hopefully by now you understand what Transformer models are, and why they are so popular and impactful. In this section I will introduce a catalog of the most important Transformer models that have been developed to this day. I will categorize each model according to the following properties: Pretraining Architecture, Pretraining Task, Compression, Application, Year, and Number of Parameters. Let's briefly define each of those:

#### 2.1.1 Pretraining Architecture

We described the Transformer architecture as being made up of an Encoder and a Decoder, and that is true for the original Transformer. However, since then, different advances have been made that have revealed that in some cases it is beneficial to use only the encoder, only the decoder, or both.

**Encoder Pretraining** These models, which are also called bi-directional or auto-encoding, only use the encoder during pretraining, which is usually accomplished by masking words in the input sentence and training the model to reconstruct. At each stage during pretraining, attention layers can access all the input words. This family of models are most useful for tasks that require understanding complete sentences such as sentence classification or extractive question answering.

**Decoder Pretraining** Decoder models, often called auto-regressive, use only the decoder during a pretraining that is usually designed so the model is forced to predict the next word. The attention layers can only access the words positioned before a given word in the sentence. They are best suited for tasks involving text generation.

**Transformer (Encoder-Decoder) Pretraining** Encoder-decoder models, also called sequence-to-sequence, use both parts of the Transformer architecture. Attention layers of the encoder can access all the words in the input, while those of the decoder can only access the words positioned before a given word in the input. The pretraining can be done using the objectives of encoder or decoder models, but usually involves something a bit more complex. These models are best suited for tasks revolving around generating new sentences depending on a given input, such as summarization, translation, or generative question answering.

# 2.1.2 Pretraining Task

When training a model we need to define a task for the model to learn on. Some of the typical tasks, such as predicting the next word or learning to reconstruct masked words were already mentioned above. "Pre-trained Models for Natural Language Processing: A Survey"[10] includes a pretty comprehensive taxonomy of pretraining tasks, all of which can be considered self-supervised:

- 1. Language Modeling (LM): Predict next token (in the case of unidirectional LM) or previous and next token (in the case of bidirectional LM)
- 2. Masked Language Modeling (MLM): mask out some tokens from the input sentences and then trains the model to predict the masked tokens by the rest of the tokens
- 3. **Permuted Language Modeling (PLM):** same as LM but on a random permutation of input sequences. A permutation is randomly sampled from all possible permutations. Then some of the tokens are chosen as the target, and the model is trained to predict these targets.
- 4. **Denoising Autoencoder (DAE):** take a partially corrupted input (e.g. Randomly sampling tokens from the input and replacing them with "[MASK]" elements. randomly deleting tokens from the input, or shuffling sentences in random order) and aim to recover the original undistorted input.
- 5. Contrastive Learning (CTL): A score function for text pairs is learned by assuming some observed pairs of text that are more semantically similar than randomly sampled text. It includes:
  - **Deep InfoMax (DIM):** maximize mutual information between an image representation and local regions of the image;

- Replaced Token Detection (RTD): predict whether a token is replaced given its surroundings;
- Next Sentence Prediction (NSP): train the model to distinguish whether two input sentences are continuous segments from the training corpus; and
- Sentence Order Prediction (SOP): Similar to NSP, but uses two consecutive segments as positive examples, and the same segments but with their order swapped as negative examples

### 2.1.3 Application

Here we will note what are the main practical applications of the Transformer model. Most of these applications will be in the language domain (e.g. question answering, sentiment analysis, or entity recognition). However, as mentioned before, some Transformer models have also found applications well beyond NLP and are also included in the catalog.

#### 2.2 Catalog table

Figure 5 is a screenshot of a large table where I have tabulated all the models. If you are interested in the table, access it directly in  $^{18}$ 

# 2.3 Family Tree

The diagram in figure 6 is a simple view that highlights the different families of transformers and how they relate to each other.

#### 2.4 Chronological timeline

Another interesting perspective of the catalog is to see it as a chronological timeline. In Figure 7 you will find all the transformers in the catalog sorted by their date of publication. In this first visualization, the Y-axis is only used to cluster transformers of related heritage/family.

In Figure 8, the Y-axis represents model size in millions of parameters. You won't be able to see all the models in the catalog since many fall right on the same time and size, so please refer to the previous image for that.

# 2.5 Catalog List

Finally, here is the full list view that might be easier to follow along in some cases:

#### 2.5.1 ALBERT

- **Reference:**<sup>19</sup>[11]
- Family: BERT
- Pretraining Architecture: Encoder
- Pretraining Task: MLM/NSP
- Extension: Compressed version of BERT using parameter sharing, which is much more efficient given the same number of parameters
- Application: Same as BERT
- Date (of first known publication): 09/2019
- Num. Params:b12M, Large = 18M, XLarge = 60M\*
- Corpus:Same as BERT
- Lab: Google

#### 2.5.2 AlphaFold

• **Reference:**<sup>20</sup>[12]

<sup>&</sup>lt;sup>18</sup>https://docs.google.com/spreadsheets/d/11tyrAB6BL29cOv2fSpNQnnq2vbX8UrH147d7FkIf6t4

<sup>&</sup>lt;sup>19</sup>https://huggingface.co/docs/transformers/model\_doc/albert

<sup>&</sup>lt;sup>20</sup>https://www.deepmind.com/publications/highly-accurate-protein-structure-prediction-with-alphafold

|  
  | -   |  
   |   
  |  |   | Data inf first  |   
  |  |   
   |
|---
--|---
--
--|--|--|---
---	--
  |   | Building 1 - 199 - 1 - 199 - 1 - 199   
   | Prezaksing  
  |  |   | known   | -   
  | A  |   
   |
| Name of Madel  
  | Farmy   | Prenaming Anothicitary (Lineader Dicedents-O)  
   | 1244  
  | LUBRER   | Approximin  | publication)  | Base = 1250,  
  | Corpor   | Lib   
   |
| ALBERT   
  | BERT  | Grouder  
   | M.M.NSP   
  | Compressed version of BERT using parameter sharing, which is much more efficient given the same<br>number of parameters  | Same as BERT  | 09(201  | Largo = 50M,<br>XLargo = 60M  
  | Same as BERT   | Google  
   |
|  
  |   |  
   |   
  | The original Alphafold used a BERT-style transformer. The details of Alphafold's Transformer are not   |   |   |   
  |  |   
   |
| AlphaFold  
  | 9512)-Transformer   | Enceder  
   | prediction  
  | (see this blog port)   | Protein faiding   | 071202  | 210   
  | 170,000 proteins from a public repository of protein sequences and structures  | Deepmind  
   |
|  
  |   |  
   |   
  | These models do not introduce neverilies at the architecture/pretraining level and they are based on<br>GPT-3 but rather focuses on how to improve alignment through fine-tuning and prompting. Note that  |   |   |   
  |  |   
   |
|  
  |   |  
   |   
  | the Arthropic Assistant includes several models optimized for different tasks. Latest versions of this<br>and forces on the basedite of BLHE.  | Different models with different annihilations from carsetal dialog in   |   |   
  | 4508 Internet from Marcel Common Crewit and Books. They also mental seasonal Dialogue Professors distances for   |   
   |
| Anthropic Assistant (see also)   
  | GPT   | Decoder  
   | LM  
  |  | code assistant.   | 121202  | 104115 520  
  | the RLHF training  | Anthropic   
   |
| BART   
  | GPT for Decoder   | Encoder/Decoder  
   | DAE   
  | It can be seen as a generalization of BERT and GPT in that I combines deal from both in the<br>encoder and decoder   | Mostly test generation but also same test understanding tasks   | 10(201  | BERT  
  | Same as RuleERTs (1605b of news, books, Morkes,<br>and web text)   | Facebook  
   |
| GERT   
  | BERT  | Encoder  
   | MMNSP   
  |  | General Language Understanding and Question Answering. Many other language applications followed  | 10(201  | Bass = 190M,<br>Large = 340M  
  | Toronto Book Corpus and Welpedia (3.38 Tokena)   | Google  
   |
|  
  |   | Encoder AND Encoder/Decoder (BigBird is mostly a way to  
   |   
  | Big Bird can oxdend other architectures such as BERT, Pegasus, or RoBERTs by using a sparse  | Destroited and a first test second and and being the  |   | Depends on  
  |  |   
   |
| Gig Gird   
  |   | Encoder-only as wells as Encoder Decoder architecture)   
   | MLM   
  | sequences  | e.g. in genomics  | 071202  | architecture  
  | Books, CC-News, Stories and Wikipedia  | Geogla  
   |
|  
  |   |  
   |   
  | Biendefligt 3 is based on a pre-trained OPT. It adds features needed for a dialog agent such as<br>long-term memory or the ability to search the internet. It is also fine-tuned for some specific tasks   |   |   |   
  |  |   
   |
| Glanderfliet, 3  
  | GPT   | Decoder  
   | LM  
  | givenhumen feedback on them.   | same as GPT-3   | 06/202  | 2 1758  
  | 1808 tokens = RoBERTs + the Pile + PushShilt.io Reddt  | Facebook  
   |
| BLOOM  
  | GPT   | Decoder  
   | LM  
  | Main difference to GPT-3 is that it uses full attention instead of sparse attention  | Same as GPT-3   | 07/202  | 1768  
  | 3668 tokens (1.5 TB of lost data) multilingual dataset   | Sciencel-lugging/ace  
   |
|  
  |   |  
   |   
  | ChatGPT takes a GPT3.5 (also GPT3 Devind-003) pretrained readel and uses RU-F to fineture the<br>model mostly like described in instructGPT but with slight differences in the data collection. ChatGPT  |   |   |   
  |  |   
   |
| CharGPT  
  | CRT   | Decoder  
   |   
  | is also more than a model since it includes extensions for Memory Store and retrieval similar to<br>Disorderitor's   | Dialog anests   | 10/202  | Same as GPT1  
  | Same as GPT3 a detends reparated for B116  | Oraciti   
   |
|  
  |   |  
   |   
  | Same as Gopher but with optimizations to reduce model size and therefore training interance time   |   |   |   
  | Massive Test (2.35 billion documents, or about 10.5 TB of test including Massive Web, Books, Github, Neves, C4,  |   
   |
| chirchila  
  | GPT   | Decoder  
   | LM<br>oredict which a   
  | with equal or superior performance   | Same as Gopher/GPT3   | 03/202  | 2 756   
  | and Wikipedia.   | Deepmind  
   |
|  
  |   |  
   | the N + N   
  |  |   |   |   
  |  |   
   |
|  
  |   |  
   | (image, text)   
  |  |   |   |   
  |  |   
   |
|  
  | CLIP (Also using  |  
   | across a batch  
  |  |   |   |   
  |  |   
   |
|  
  | vanila tanaformar   |  
   | occurred  
  |  |   |   |   
  |  |   
   |
| SME  
  | for skep  | L'ALIGN  
   |   
  | Contention Preside and VIT for the Vecal encoding with Interdometer for the Telebal encoder  | Multimodal language model with the ability to do structured   |   |   
  | min (vieumagenes) - ess minum euromage para  | openno  
   |
| CMD  
  | HTML  | Decoder  
   | Causality-mask<br>ed LMs  
  | This is somewhat similar to HTML in its use of structured training data. However, it is a different<br>architecture and uses causal masking  | prompting   | 01/202  | 13B (Arrport)   
  | CC-News. English Wikipedia   | Facebook  
   |
|  
  |   |  
   |   
  |  |   |   |   
  | 143 GB of test including: Wikipedia (En, De, Ea, Fr), Project Gutenberg, 45 subreddite, OpenWebTest2, Amazon   |   
   |
|  
  |   |  
   |   
  | model can generate test conditioned on control codes that specify domain, style,   |   |   |   
  | Heveres, subget and UV data from HMF, guildon-interer part from LLD, and the<br>MRDA shared balk), which includes the Stanford Question Answering Dataest, NewsQA, TrivisQA, SearchQA,   |   
   |
| CTRL   
  |   | Decoder  
   | Caption   
  | topics, dates, entities, relationships between entities, protiponts, and task-related behavior.<br>A differential variational auto-encoder is used to learn the visual codebook. The travelorner is a  | Controllable test generation  | 054201  | 1.638   
  | HarpotQA, and Natural Questions  | Salentorce  
   |
| DALL-E   
  | GPT   | Decoder  
   | prediction  
  | variation of GPT-3   | Text to image   | 1/202   | 120   
  | 250 million tool-images pairs from the internet  | OpenAl  
   |
| DALL-E-2   
  | CLIP, GLIDE   | Enceder/Decoder  
   | prediction  
  | Combines CLIP encoder and Diffusion decoder similar to GLIDE   | Text to image   | 41202   | 3.58  
  | Combination of the DMLL-E and CLIP datasets  | OpenAl  
   |
|  
  | GPT, Control<br>Transformers" (not  |  
   |   
  |  |   |   |   
  |  |   
   |
|  
  | per so a family but grouping have three   |  
   |   
  |  |   |   |   
  |  |   
   |
|  
  | transformers that by  |  
   |   
  |  |   |   |   
  |  |   
   |
| Desision Transformers  
  | general control,<br>Di Jive trates  | Decoder  
   | Next action   
  | Decision transformers use a GPT architecture and extend it by encoding trajectories in a way that  | General R1 Interferoment Institute and a  |   | Same - Com  
  | Different corrus for different exceptions  | Google UC<br>BeckelouiD 10  
   |
| OwkGPT   
  | GPT   | Decoder  
   | LM  
  | GPT-2 architecture trained on claing data  | Text generation in dialog settings  | 10(201  | 1.50  
  | 143M Redit conversions   | Microsoft   
   |
| OwnEDT   
  | DEPT  | Eventer  
   | M MANTO   
  | Compressed version of BERT using distillation, which is much more efficient given the same number of provident   | Garna as DEDT   |   | CEM .   
  | Gama as BEDT   | likesinete -  
   |
|   |   |  | -ALTER (DIP  |  | and an address of the second s  | 10(201  | Lip to 20s   |  | - Stadiega  |
|  
  |   |  
   |   
  |  |   |   | reduction in<br>pataneology   
  |  |   
   |
| TRANSCO  
  | BART  | ExcederDecoder   
   | ONE   
  | Adds supplication and distillation is a BART model to improve performance and model vice   | Text accession and understanding  | 0,000   | compared to   
  | CANUDM XSUM ELIS WATTH EX-BO (-1911)   | Amazon  
   |
| and server a  |   |  |  | the second  |   | 0.0200  | Base = 190M.   | And a second sec   |   |
| ELECTRA  
  |   | Encoder  
   | RTD   
  | Uses BERT for Encoder architecture, but stacks and accrecates two of them for last evolventies   | Some as BERT<br>Knowledge intensive related tasks that might benefit from incontents  | 03/202  | Large = 33064   
  | Same as BERT except for Large with is name as XLNet  | Stantset/Google<br>Various Chinase  
   |
| CRNC   
  | BERT  | Encoder  
   | MLM .   
  | This architecture could be understood as BERT for test + knowledge graphs  | graphs or entities such as entity recognition   | 05/201  | 15484   
  | English Wikipedia + Wikidata for entitizes (note that they initialize model to original BERT parameter values  | institutions  
   |
|  
  |   |  
   | Log likelihood<br>of test given   
  |  |   |   |   
  | MultiWadal MasalvelHob (H2W): 185 million images and 192 GB text + a number of text paired with image  |   
   |
| Ramingo  
  | Chinchile   | Decoder  
   | spre visual<br>inpet  
  | It uses a frazer testual language model (ike Chinchila) conditioned on the visual representation,<br>which is encoded from a Normalizer-Free ResNet.   | Tood to image   | 41202   | 100 (argent)  
  | datasets: ALIGN + LTP (Long Ted & Image Paint) = 312 million images, and VTP (Video & Ted Paint) = 27<br>million short videos (approximately 22 seconds on average)  | Deepmind  
   |
|  
  | "Control<br>Transforment" local   |  
   |   
  |  |   |   |   
  |  |   
   |
|  
  | per se a family, but  |  
   |   
  |  |   |   |   
  |  |   
   |
|  
  | transformers that try   |  
   | MLM (where  
  |  |   |   |   
  |  |   
   |
|  
  | general cantrol,  |  
   | either text or  
  | The standard decoder-only transformer architecture is preceded by an embedding layer that can  | Gato presents a generalizable agent that can be used beyond lood to   |   |   
  | 1.5T takens including standard test (e.g. MassiveTest), vision (e.g. AUGN), and simulation environments (e.g.  |   
   |
| Ginto  
  | RL-like, tasks)   | Decoder  
   | agent actions)  
  | embed test and images, plus add position encodings to add spatial information when applicable.   | tasks such as playing Atari or controlling a robot arm.   | 5/202   | 1.20 1.20 1.27 across 64  
  | ALE Atari, or RGB Stacking Real Robot)   | Deepmind  
   |
|  
  |   |  
   |   
  | Gi all introduces a blocks of 64 Events in increase parameter court and executivelog properties  |   |   | esperts, but  
  |  |   
   |
|   |   | B  |  | in a somewhat standard decoder-only. Terreformer architecture. Only two esperts get activated at a   |   |   | activated for  | A REVISION INCOMENTATION OF A DESCRIPTION OF A DESCRIPTIO   |   |
| Sec. Ma  
  | Incention   | L'eccew  
   | LN  
  | time per teen, which makes the model and more encent in training and interance.  | General language modeling   | 121210  | 3.50 diffusion  
  | 1.6.1 tokens including web pages thered by welpecia and books for quality  | Googa   
   |
|  
  |   |  
   |   
  |  |   |   | model (2.38 for visual  
  |  |   
   |
|  
  |   |  
   |   
  | GLDE can be seen as an extension of the ADM (Ablated Diffusion Model) by the same authors.<br>Unserver, ADM is not ner to a transformer architecture although it does teaching one is some of the  |   |   | encoding, 1.28  
  |  |   
   |
| 0.00   
  | Olivering module  | Formeter   
   | Caption   
  | configurations the authors use. Given that AOM is by the same authors and was quickly followed up<br>to QUID_ Utility is for to sensitive QUID_ as the first of its first.   | Text in image   | 10000   | 1.58 for model  
  | Energy on Dall 1-5   | Quanti  
   |
| -  
  | Constant Fielder  |  
   | image   
  |  | in a rage   | 12/200  | - na opeander g   
  | AND MILLION AND AND AND AND AND AND AND AND AND AN   | openny  
   |
| Globel Context WT  
  | VIT   | Enceder  
   | Classification  
  | hierarchical WT architecture consisting of local and global self-attantics modules   | Image (object detection, image classification)  | 06/202  | 2 9644  
  | bragenet-IK and other task dependent datasets  | Nividia   
   |
|  
  |   |  
   |   
  |  |   |   |   
  | Information Total (1.10) follows does strengths, or advocal 10.7. The of heat total officer Managine Month Photons, 7100-05, houses, 714   |   
   |
| Gapter   
  | GPT   | Decoder  
   | LM  
  | same as or 1-2 bet use response interact of Layereant and readow pasticing remaining about the that absolute   | Mostly Language Madeling and NLU, but also extensible like GPT  | 121202  | 2008  
  | and Wikipedia.   | Deepreind   
   |
Gapter	GPT Geatur	Decoder Decoder	LM LM	Same as CMP-12 but can reason instance or Lagranger and nearest parameters encouring memory abacitation Coppercible is based on Copper but adds a step same RLI-PP (Reinforcement Learning from Human Professionals) to learn whether on only an encourse is obtained but allow control	Mostly Language Modeling and N.U., but also adsemble like GPT Diston posterors. GBA: senseral because conversion tasks	12/202	2008	energie energi	Despreind
Gopher GopherCite	GPT Gepher	Decoder Decoder		Salma as OF For Declars Report Instance of Layerson and instance parameters parameters encoding interaction administration of the Declarson of Cogher Full adds a step assing RELPE (Roleforcement Learning from Human Profesence) to learn whether not only a response is placable but also sugported	Mostly Language Modeling and NLU, but also extensible like GPT Dialog systems, QEA, persent language generation basis Tool generation, but adaptable to many other NLP tasks when five	121202	2008	neueron an Cashar Felix and a second and a second s	Deepmind
Gupter GopterCite GPT	GPT Gegher GPT	Decoder Decoder Decoder		and is an of the second and experiment instance of copyregation and intervery paradional instances of the second on Gaptive bot and a range uning (REVP (Paintersament) Learning from Human Performance) to summit whether not entry by negociaria is placed but also suggestive of the second on the the OPT architecture (as in the remaining the model).	Mostly Language Modeling and NLU, but also extensible like GPT Dialog systems, OAA, general language generation tasks Tool generation, but adaptable to many other NLP tasks when fine tande.	121202 031202 062201	1 2000 2 2000 1 11714	and Wilspeck. On the International Control of the Net o	Deepmind Deepmind OpenAl
Gapter GapterClas GPT GPT2	GPT Gepher GPT GPT	Decoder Decoder Decoder	ui ui ui ui	Analysis of the second and the second	Mostly Language Modeling and N.U., but also solveable like GPT Dialog systems, GAA, general lenguage generation tasks Too generation, but adaptable to many other NLP tasks when fine lands. Too generation, but adaptable to many other NLP tasks when fine lands.	12/202 03/202 06/201 02/201	2000 2000 11704 1150	ent Mitiged a. Sama in a capacity fail and a particle dataset provided in the IR-IP process failed and the second second second second second second second take question of a second gradual REAC Question. Territories was been provided and the second second second second second second second fail and the second second fail and the second second fail and the second	Deepmind Deepmind OpenAJ OpenAJ
Gapher GeptenCite GPT GPT2	GPT Gegher GPT GPT	Decoder Decoder Decoder Decoder		Sense. The sense of the sense of the sense of the sense is a sense of the sense of	Mostly Language Modeling and NLU, but also extensible like GPT. Dialog spearster, GLA, general language per-antice taxia. Tool spearating but adaptable to range pether NLP tools when fiss faced. Tool perioditise, but adaptable to range other NLP tools when fiss food periodics. But adaptable to range other NLP tools when fiss frong of applications. But has now titte base used for a lange writing basevarias. But man ow titte bases used for a lange	12/200 03/200 06/201 02/201	2 2000 2 2000 0 11704 0 1.50	And Hillingwide Start Hillingwide Hillingwide (Provide Start) and Anton Provide Start (Samiran Start) Hillingwide (Provide Start) and Start (Samiran Start) Hillingwide (Provide Start) (Samiran Start) Hillingwide (Provide Start) (Samiran Start) Hillingwide (Provide Start) (Samiran Start) Hillingwide (Provide Start) (Samiran Start) Hillingwide (Provide Sta	Despreind Despreind OpenAJ OpenAJ
Gapter GapterCan Gett Getta Getta	GPT Gegher GPT GPT	Decoder Decoder Decoder Decoder Decoder		and an and an operation of the second seco	Martly Language Modeling and NLU, for allow extended like GPT Dridg systems, DAA, personal language generation takes that approximate, has industable to many other NAP basis where free hands. The analyzable to many other NAP basis where free hands in the analyzable to many other NAP basis where free hands. The analyzable to many other NAP basis where free hands in the analyzable to many other NAP basis where free weight and generation. But has over time been used for a large and analyge and analyge periodian. But also many of analog periodian.	12/200 03/200 06/201 02/201 05/200	1 2008 2 2008 8 117M 8 1.50 1.50 1.50	Let Rillied. Let Rillied. Unspectra Physics and Categoria Internation 7 to SUP process Unspectra Physics on Disadorps data Expected Transmis on some tain querk, datases internation 944, Mort Categoria. Transmission gauge (31 db) 505 GPT Worth of datases is avoid by graving of hour at bacterist in taine 3 disagregation. - 2680 Internationality Categoria (1980), Morthad (1980, Bacteri (1980, Bacteri (1980, avoid 1980, and Magade 1981)	Deepmind Deepmind OpenAl OpenAl OpenAl
Gaster GasterCite Gasta Gasta Gasta Gasta Gaster	GPT Gepher GPT GPT GPT	Decoder Decoder Decoder Decoder Decoder Decoder Decoder		The second seco	Netly Larguage Modeling and NUU, but days unsmaller like QPT Child in generations. GML prevent language parentize tasks and parametisk. In bindings in honry of text BD and whet for the approxedia is bindings in honry of text BD and whet for the approxedia is bindings in honry of text BD and whet for the approxedia is bindings in the only prevent by a large and approxedia is bindings in the only prevent by a large and approxedia is bindings in the only prevent of the large and approxediance, but aligned in the only prevent of the large and approxediance, but aligned in the only prevent of the large and approxediance, but aligned in the only prevent of the large and and approxediance. But aligned in the only open NAP basis when fits and approxediance.	12200 03200 04201 02201 02201 05200	2000 2000 11776 11776 11786 11786 11786 11786 11786 11786 11786 2178 (0)	ear Mitodes. I ear Mitodes. The and Gapter Las and Mitodes and percendent if a full of the mean and the second the second theory and a mean second second and the second and the second second the second full ear and any ear of Gis. 100 CPT. Which the advance is executed by another, and the at last 25 Mitodes and any ear of Gis. 100 CPT. Which the advance is executed by another, and the at last 25 Mitodes and the second s	Deeprind Deeprind OpenN OpenN OpenN EisetherN
Gashar GasharChe Gett Getta Getta Gettaleg	GPT Geptwr GPT GPT GPT	Dender Dender Dender Dender Dender		means. We have a second secon	Metry Lenguage Molding and KUJ, and an on-orderable line GPT "Data goalens," GML general impaging personality limits and the second second second second second second second the second second second second second second second second impages and the second second second second second second responses of the second second second second second second second responses of the second second second second second second second second responses of the second sec	12200 03200 06201 02201 02201 05200	2 2008 2 2008 0 11784 0 1.558 1.558 1.558 2.718 (PL)	Let Missian Let Missian Disponse for provide and the dense for example of the GLE provides of the second wave, where Disponse provides and the dense for a second second second second second second the dense of the dense of the dense of the dense of the second seco	Deepmind Deepmind OpenAl OpenAl DeenAl
Souther Gaptin-Cale Gaptin-Cale Gapting Gaptin	CPT Cepher CPT CPT CPT CPT CPT CPT CPT	Dender Dender Dender Dender Dender Dender Dender		makes the second secon	Methy Length Modeling and NUU, but does not watch the GPT Child og begreven. GAL y Lener Managemen generation tests Test generations. The strength in terms of the HT bases where the managements. The strength is the mark of the HT bases where the test generation. The strength is the test of the strength is defined and the strength in terms of the HT bases where the strength of and generation. The strength is the strength renge of adjustments. The strength is the strength renge of adjustment in the strength and the strength renge of adjustment in the strength and the strength renge of adjustment in the strength and the strength and the strength and the strength and the strength and the strength and strength and the strength and the strength and the strength and strength and the strength and the strength and the strength and strength and the strength and the strength and the strength and strength and the strength	12000 030200 06200 05200 05200 030200 030200 04202	2 2008 1 11714 1 1250 1 1250 1 1250 1 22 2006	ear Mitoda	Desprind Desprind OpenAl OpenAl DesrAl EleuthonAl EleuthonAl
Gester GesterCin Gett Gett Gett Gett Gett Gett Nuck 200 Gett Nuck 200	GPT     Geptwr     GPT     GPT     GPT     GPT     GPT     GPT     GPT	Dender Dender Dender Dender Dender Dender Dender		makes. Market and the second	Media Language Makeling and KLU, Lan Kan Kanakita Kito CPT Child parketini. CAL Anna Kanakita Kito CPT Child parketing the sector barry Tang an entropy. In a language material transfer and the sector of the secto	13200 03200 04201 02201 04202 04202 04202	2 2938 2 2938 4 11734 4 1.528 3 135 18 1.528 2.738 (RL) 2 2005 2 2005	ear Mitoda's ear Mitoda's Hampsong Carlos (La Service) parameter i la Service) Processo Hampsong Carlos (La Service) (La Service) Processo (La Service) (La Service) Hampsong La Service) (La Service) (La Service) (La Service) (La Service) (La Service) (La Service) Hampsong La Service) (La Service) Hampsong La Service) (La Service	Desprind Desprind OpenAI OpenAI Desprind Desprind Desprind Desprind Desprind Desprind
Seater         SepterCla           SepterCla         SepterCla	арт аврыг арт арт арт арт арт арт арт арт	Desater Creater Create		makes makes where the second sec	More is sequence that the or to U. I. A rate is submitted in GOT to graph setup. (A), the sequence process that the OT many periods to the sequence process that the origination of	12200 03200 062201 05200 03200 042322 10200	2 2838 2 2838 4 11734 4 1.528 2 155 B 1.528 2 778 (RL) 2 2005 2 17558	eer Maaks. I een Maaks. I een Geren van aan die de een eerste in te huit een op	Despreind Despreind OpenAl OpenAl DesuthenAl DisuthenAl OpenAl
Seate           SepterCite           SPT           SPT3	GPT GgPar GPT GPT GPT GPT GPT GPT GPT	Desater Constar Onstar Desater Desater Desater Desater Desater		makes. Market and the second	Much propagate sharing on U.S. J. on an exhand in OPT Congo systems. Of A systems in the steps provide the steps The product of the steps provide the steps of the step of the	15200 05201 05201 05201 05200 05200 05200 05200 04202 10200 01200	2 2438 2 2438 4 159 4 159 4 159 4 159 2 278 pt.] 2 208 2 208 2 1558 2 1558 2 1558	ear Market. Ear Market. Hann Golgen Jack and Hannes Jacobine 1 for BUT process Hanneson Conference and Hannes Annotation 1 for Annotation and Hannes Annotation Hanneson Annotation 1 for Annotation 1 for Annotation 1 for Annotation Hanneson Annotation 1 for Annotation 1 for Annotation 1 for Annotation Hanneson Annotation 1 for Annotation 1 for Annotation 1 for Annotation 1 for Annotation Hanneson Annotation 1 for Annotation 1 for Annotation 2 Jacobine 2 Jacobine 2 Balance Phe J 2 Annotation 1 for Annotation 2 Jacobine 2 Jacobine 2 Jacobine 2 Balance Phe J 2 Annotation 1 for Annotation 2 Jacobine 2 Jac	Despreind Despreind OpenAl OpenAl DestManAl DestManAl OpenAl OpenAl
Sector Sector Clin 2073 2073 2073 2073 2073 2073 2073 2073	орт орт орт орт орт орт орт орт	Deader Deader Oracler Deader Deader Deader Oracler Deader Deader		makes makes where the second sec	Medy Language Underling on KUU, La van sonkreide In OPT Top person (Kan Server and Serv	15200 05201 05201 05201 05200 05200 05200 05200 04202 04202 04202 04202 04202 04202 04202 04202 04202 04202 04202 04202 04202 04202 04202 04202 04200 0400000000	2608           2008           1178           158           158           278 PLj           208           158           158           158           158           158           158           208           208           1758           208           1758           208	ear Maydes. I ear Maydes. The and Gapter Las and the densel prevented in the ULE process. Theory and the process of the Control Process of the control take expected in the annual theory and the process of the Control Process of the control take expected takes and the control Process of the Control Process of the control take expected takes and the control Process of the Control Process of the control take and the control Process of the Control Process of the Control Proc	Despreind Despreind OpenAl OpenAl DestManNi DestManNi OpenAl OpenAl OpenAl OpenAl
State           State </td <td>GPT         Gapher           GPT         GPT           GPT         GPT</td> <td>Conter Co</td> <td>LM LM LM LM LM LM LM LM LM LM LM LM LM L</td> <td>makes. Market and the second second</td> <td>Made Languages Massing and KLU, La van schundels für OFT Orse syndress. Och spranses har spranses provider har spranses har spranses har spranses har spranses provider har spranses and providers. In advances to know the Mark Mark and work for Targe providers har shrapes har spranses and the target of providers har spranses har spranses and the target of providers har spranses har har spranses har spranses and the target of providers of providers har spranses har spranses and the target of providers har spranses Known and providers har spranses and the target on target of target harding and the spranses and the target on targets harding strategies that the target on target of targets harding strategies that the Gaussian spranses and the target on targets that the spranses of the targets of targets that the spranses that the target of targets that the Gaussian spranses and the target on targets that the spranses that the target of targets that the spranses the spranses the spranses that the spranses that the spranses the spranses the spranses that the spranses</td> <td>13200 00201 00201 00201 00200 00200 00200 00200 00200 00200 01200 01200 01200 01200 01200 01200 01200</td> <td>2 2008 2 2008 4 17784 4 158 4 158 4 158 2 708 2 708</td> <td>ear Market. Ear Market. Here in Gapera Jack and Here and personal in the GaP process Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company of the Company Hampson Development of the Company of the Co</td> <td>Desprind Desprind OpenAl OpenAl DestherNi DestherNi OpenAl OpenAl Facebook Googe</td>	GPT         Gapher           GPT         GPT	Conter Co	LM LM LM LM LM LM LM LM LM LM LM LM LM L	makes. Market and the second	Made Languages Massing and KLU, La van schundels für OFT Orse syndress. Och spranses har spranses provider har spranses har spranses har spranses har spranses provider har spranses and providers. In advances to know the Mark Mark and work for Targe providers har shrapes har spranses and the target of providers har spranses har spranses and the target of providers har spranses har har spranses har spranses and the target of providers of providers har spranses har spranses and the target of providers har spranses Known and providers har spranses and the target on target of target harding and the spranses and the target on targets harding strategies that the target on target of targets harding strategies that the Gaussian spranses and the target on targets that the spranses of the targets of targets that the spranses that the target of targets that the Gaussian spranses and the target on targets that the spranses that the target of targets that the spranses the spranses the spranses that the spranses that the spranses the spranses the spranses that the spranses	13200 00201 00201 00201 00200 00200 00200 00200 00200 00200 01200 01200 01200 01200 01200 01200 01200	2 2008 2 2008 4 17784 4 158 4 158 4 158 2 708 2 708	ear Market. Ear Market. Here in Gapera Jack and Here and personal in the GaP process Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company Hampson Development of the Company of the Company of the Company of the Company Hampson Development of the Company of the Co	Desprind Desprind OpenAl OpenAl DestherNi DestherNi OpenAl OpenAl Facebook Googe
0x84x 0x84xC4a 0752 0753 07555 0755 0755 0755 0755 0755 0755 0755 0755	GPT Gepter GPT GPT GPT GPT GPT GPT GPT GPT	Conter Co	LM LM LM LM LM LM LM LM LM LM LM LM LM L	makes makes which is a second se	Medy Language Underling on KUU, La van senarheits in GPT Tell generation, Ka senarheits in GPT and an eleven the Tell generation, Ka senarheits in some yet her KP med and en eleven methods and an eleven to senarh in the senarheits of tell generation. Tell generation, Ka senarheits in the senarheits of tell generation, the senarheits in stationage that KP med and en eleven tell generation, Ka senarheits and an eleven the tell generation, the senarheits in stationage that KP media and en factor tell generation, Ka senarheits and tell generation the tell generation, Ka senarheits and tell generation. Tell generation, Ka senarheits and tell generation. Tell generation, Ka senarheits and tell senarheits Tell senare. Tell senare.	19200 066201 066201 065200 065200 065200 064292 106200 016200 016200 016200 016200	2 2003 2 2003 4 17784 4 1528 1 528 1 528 2 2005 2 2005 2 3005 2 3005	eer Maake. een Maake. Teen Garper van eer die Gewaard personiel is 'n Uie Process Teasporter Property op Handler Stander, Teaster Van een van teen eerde kernen Hander van die gewaard die Stander Van Verster Van eerde van de stander die stander Kerne zahle. Calif Beeren kan die Stander Van Verster Van Verster Van eerde van die stander die stander Verster zahle. Teaster van die stander van die stander van die stander van die stander van die stander Maake van die stander van die stander van die stander van die stander Teaster van die stander van die stander van die stander van die stander Teaster van die stander van die stander van die stander van die stander Teaster van die stander van die stander van die stander van die stander Teaster van die stander van die stander van die stander van die stander Teaster van die stander van die van die stander van die stander van die stander Teaster van die stander van die van die stander van die stander van die stander Teaster van die stander van die stander Teaster van die stander van die stander van die van die stander van die stan	Desprind Desprind OpenAl OpenAl DestherM DestherM OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl
0x84x 0x72 072 072 073 073 073 073 073 073 073 073 073 073	GPT Gigher GPT GPT GPT GPT GPT GPT GPT GPT	Denter Denter Obsard Denter De	LM LM LM LM LM LM LM LM LM LM LM LM LM L	makes. Market and the second	Made (a sequence) standing on U.S. (a set on schedule the OFT Consequence). Call sequence and the set of the Set of Set o	12000 03000 06001 05000 04002 10000 04002 10000 01000 07100 07000000	2 2003           2 2003           4 11734           4 152           1 153           1 153           1 153           1 153           1 153           1 153           1 153           1 153           1 153           1 153           2 2005           6000           2 30           7530 (Jumbol)           7530 (Jumbol)           7530 (Jumbol)	ear Market. I ear Market in Section 2 and	Desprind Desprind OpenAl OpenAl DestAnd DestAnd DestAnd OpenAl OpenAl OpenAl OpenAl Al Cogije
State           State           Get1           Get12           Get13           Get14           Get15           Get14           Get15           Get15           Get16           Get17           Get17           Get18           Get19           Get19           Get19           Get19           Get19           Get19 <td>GPT Gaghar GPT GPT GPT GPT GPT GPT GPT The GPL Dilation GPT The GPL Dilation GPT The GPL Dilation</td> <td>Conter Co</td> <td>LM LM LM LM LM LM LM LM LM LM LM LM LM L</td> <td>makes makes submitted by the second sec</td> <td>Med (a sequel shading on KU), a car is submitted in GOT Top periods (A), a sequel shading periods the SU Top periods (A), a sequel shading periods the SU Top periods (A) and a sequel shading a sequel to the sequel shading a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU SU (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) SU (A) and (A) SU (A) and (A)</td> <td>12000 05000000</td> <td>2000           2000           2000           1500           1500           1500           2000</td> <td>eer Maake. een Maake. Hen Geren kaar eerste fer deen geveende 14 e. G.U.F. provide Transporter for Young op Inductorput Maake. Expenden 17 providege moet take quelet is areas Henry and any op 16 Ge. 100 GPT. Yee'te fait Maake it week by analety of total affected at later Young and the second op 16 Ge. 100 GPT. Yee'te fait Maake it week by analety of total affected at later Young and the second op 16 Ge. 100 GPT. Yee'te fait Maake it week by analety of total affected at later Young and the second op 16 Ge. 100 GPT. Yee'te fait Maake it week by analety of total affected at later The JDE dates mersely the second op 16 Ge. 100 GPT. The second op 16 Ge. 100 GPT. The second op a second op dates in The JDE dates mersely the second op a second op dates in The JDE dates mersely. The JDE dates mersely the second op a second op dates in JDE dates mersely. The JDE dates mersely the second op a second op dates in the second op 10 Ge. 100 GPT. The second of JDE mersels with the second op and the second op a second op dates in JDE dates mersels. Henry the second op 10 Ge. JDE mersels and the second op 10 Ge. JDE mersels are second op 10 Ge. JDE mersels and the second op 10 Ge. JDE mersels and the second op 10 Ge. JDE mersels are second op 10 Ge. JDE mersels and the second op 10 Ge. JDE mersels are second op 10</td> <td>Desprind Desprind OpenAl OpenAl Destendi Destendi Destendi OpenAl Facebook OpenAl Acci</td>	GPT Gaghar GPT GPT GPT GPT GPT GPT GPT The GPL Dilation GPT The GPL Dilation GPT The GPL Dilation	Conter Co	LM LM LM LM LM LM LM LM LM LM LM LM LM L	makes makes submitted by the second sec	Med (a sequel shading on KU), a car is submitted in GOT Top periods (A), a sequel shading periods the SU Top periods (A), a sequel shading periods the SU Top periods (A) and a sequel shading a sequel to the sequel shading a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU top periods (A) and a sequel shading a sequel to the SU SU (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) and (A) and (A) and (A) and (A) and (A) SU (A) and (A) SU (A) and (A) SU (A) and (A)	12000 05000000	2000           2000           2000           1500           1500           1500           2000	eer Maake. een Maake. Hen Geren kaar eerste fer deen geveende 14 e. G.U.F. provide Transporter for Young op Inductorput Maake. Expenden 17 providege moet take quelet is areas Henry and any op 16 Ge. 100 GPT. Yee'te fait Maake it week by analety of total affected at later Young and the second op 16 Ge. 100 GPT. Yee'te fait Maake it week by analety of total affected at later Young and the second op 16 Ge. 100 GPT. Yee'te fait Maake it week by analety of total affected at later Young and the second op 16 Ge. 100 GPT. Yee'te fait Maake it week by analety of total affected at later The JDE dates mersely the second op 16 Ge. 100 GPT. The second op 16 Ge. 100 GPT. The second op a second op dates in The JDE dates mersely the second op a second op dates in The JDE dates mersely. The JDE dates mersely the second op a second op dates in JDE dates mersely. The JDE dates mersely the second op a second op dates in the second op 10 Ge. 100 GPT. The second of JDE mersels with the second op and the second op a second op dates in JDE dates mersels. Henry the second op 10 Ge. JDE mersels and the second op 10 Ge. JDE mersels are second op 10 Ge. JDE mersels and the second op 10 Ge. JDE mersels and the second op 10 Ge. JDE mersels are second op 10 Ge. JDE mersels and the second op 10 Ge. JDE mersels are second op 10	Desprind Desprind OpenAl OpenAl Destendi Destendi Destendi OpenAl Facebook OpenAl Acci
Gastar         Gastar	OPT Orgher Opt	Denter Desarrow Desar	LM LM LM LM LM LM LM LM LM LM LM LM LM L	makes. Market and the second	Mad () support that by or 1(U), to an exhand its 100 T Copy system, CA, the analysis is have a local transport protection that any The porticular to alregation is have a local transport and that the protection is alregation in the system of the The analysis of the transport of the alregation is the system of the transport of the transport of the system of the system of the system of the transport of the system of the system of the system of the system of the system of the system of the system of the system of the system of the system of the system of the system Comparison system of the system of the system Comparison system of the system of the system Comparison system of the system of the system of the Comparison of the system Comparison of the system of the system of the system Comparison of the system of the system of the system of the Comparison of the system of the system of the system of the system Comparison of the system of the sy	12000 03000 02007 02007 03000 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04007 040000 04000000	2000           2000           1770/           150           150           270	ear Market. I ear Market in Section 2 and	Desprint Desprint OpenAl OpenAl DestherAl DestherAl OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl OpenAl
2x84x 2x84x2b 2x7 2x7 2x7 2x7 2x7 2x7 2x7 2x7	GPT Goglew GPT GPT GPT GPT GPT GPT GPT GPT GPT This GPD Baloff This GPD Baloff Tasuborae BAHT Tasuborae GPT	Conter Co	LM LM LM LM LM LM LM LM LM LM LM LM LM L	makes. Market and the second	Med (a sequence) whether on KUL (a set on schedule in GP) The period of the sequence period the test The period is a set of the sequence period the test the period of the sequence of the set of the sequence of the sequence is a set of the sequence of the sequence of the test of the sequence of the sequence of the sequence of the sequence of the sequence of the sequence of the sequence of the sequence of the sequence of the sequence of the test of the sequence of the sequence of the sequence of the sequence of the sequence of the sequence of the sequence of the The sequence of the sequence of the sequence of the The sequence of the sequence of the sequence of the Sequence of the sequence of the sequence of the sequence of the Sequence of the sequence of the sequence of the sequence of the Sequence of the sequence of the sequence of the sequence of the Sequence of the sequence	12000 035000 05000000	2888           2888           1731           155           155           208           208           155           208	eer Meeste een Meeste Imagewoor Proving on Househouse In the USE provide Imagewoor Proving on Househouse Imagewoor Imagewoor Imagewoor Proving Imagewoor Proving on Househouse Imagewoor Imagewoor Imagewoor Imagewoor Meeste and Imagewoor Imagewoor Imagewoor Imagewoor Imagewoor Imagewoor Meeste and Imagewoor	Desprind Desprind OperAl Destroyl Destroyl Destroyl OperAl
State         State           State         State           ST         State           State         State           State         State           State         State	OPT Coghw OPT CPT CPT CPT CPT CPT CPT CPT CCUP Diffusion CPT Tasabanar BART Tasabanar BART COPT CPT CPT CPT CPT CPT CPT CPT CPT CPT C	Conter Co	LM LM LM LM LM LM LM LM LM LM	makes. The second seco	Mad (a support sharing on U.U.), a ratio submatch in OPT Series y when CA, U.S. years in target part works that we can be a support of the strategies provide that we can be a support of the strategies and the strate of the support of the strategies in the strate of the strate of the support of the strategies in the strate of the strate of the support of the strategies in the strate of the strate of the support of the strategies in the strate of the strate of the support of the strategies in the strate of the strate of the support of the strategies in the strate of the strate of the support of the strategies in the strategies in the strategies in the support of the strategies in the strategies in the strategies in the support of the strategies in the strategies in the strategies in the Support of the strategies in the strategies in the strategies in the Support of the strategies in the strategies in the strategies in the Support of the strategies in the strategies in the strategies in the Support of the strategies in the strategies in the strategies in the Support of the strategies in the strategies in the Support of the strategies in the strategies in the strategies in the Support of the strategies in the strategies	12000 03000 04001 04000 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 04002 040000 0400000 0400000 04000000	2 3983           2 3983           4 7534           4 7534           1 528           2 7532           2 7532           2 8004           2 7532           2 8004           2 9054           2 7532           2 9054	ear Market. I ear Stages 2 and 2 an	Desprint Desprint Quest/ Quest/ Quest/ Destroyt Destroyt Quest/ Quest/ Quest/ Destroyt Destro
Satar Satar Satar Sata Sata Sata Sata Satar	GPT Goglew GPT GPT GPT GPT GPT GPT BART TC, GLP, DMain GPT BART ART GPT, DMAIN GPT	Conter Co	LM LM LM LM LM LM LM LM LM LM LM LM LM L	makes. Market and the second	Med (a support likeliho gen 14.10). Let an usuarise its 1007 Tel particular SA, second likelihood	12000 0.00000 0.00000 0.000000	2888           2888           1771           1771           1771           1771           1772           1778           2780           2780           1778           278           1778           280           1778           280           1780           280           1780           290           1780           200           1780           200           1780           200           1780           200           1780           200           1780           200           1780           1780           1780           1780           1780           1780           1780           1780           1780           1780           1780           1780           1780           1780           1780           1780           1780           1780      <	ear models.  ear models and an experimental to a model provided in the source provide  interpretent of the provide of the source provided in the source provided  interpretent of the provided interpretent of the source provided  interpretent of the source provided interpretent of the source provided  interp	Desprint Desprint Questi Questi Questi Desmonit
State         State           State         State           ST         State           State         State           State         State           State         State           State         State	GPT GPT GPT GPT GPT GPT GPT GPT	Onder Colors of	LM LM LM LM LM LM LM LM LM LM	makes. The second seco	Mad j support shafting on U.U. J and an outwards in OPT Service Service Servi	12000 03000 03000 03000 03000 04002 04002 04002 040000 04000 04000 04000 04000 04000 0400000 040000 040000 040000 0400000 040000 040000 040000 0400000 0400000 04000000	2888           2888           4724           4724           4724           725           270           270           270           270           2005	ear Market. I ear Stagest 24 and 15 and 24 percention 15 to 16 percent I happened to 26 percent 26 percention 15 to 16 percent I happened to 26 percention 26 percention 26 percent 15 percent 16 percent 26 percent 16 percent 26	Desprind Oraginal Querki Querki Destrond Destron
State         State           State         State           ST         State           ST2         State           State         State           State         State           State         State           State         State           State         State	GPT Gayw GPT	Conter Co	LM LM LM LM LM LM LM LM LM LM	makes. Makes Market	Mang Langung Handling and Kulu, Ju and Kunshandhi Kul QYT. Gang Andress, O.K. 2000, San Kulu, Juan	12000 0.00000 0.00000 0.00000 0.00000 0.000000	288           289           289           1774           1774           1774           1774           1774           1774           1775           1775           1776           1777           1778 <td>ear Market. I ear Market. I ear of Garper Jacobis and Charlow and provided in the Garper Parameter of Hermiter Barray and Hermiter Baray and Hermiter Barray and Hermiter Barray and Her</td> <td>Desprint Desprint OperAl OperAl Destrock Destrock Destrock Destrock OperAl Oper</td>	ear Market. I ear Market. I ear of Garper Jacobis and Charlow and provided in the Garper Parameter of Hermiter Barray and Hermiter Baray and Hermiter Barray and Hermiter Barray and Her	Desprint Desprint OperAl OperAl Destrock Destrock Destrock Destrock OperAl Oper
Saster Saster Sate Status Stat	Ger Gayaw Orr Ger Ger Ger Ger Ger Tasatonse Baar Tasatonse Baar Tasatonse Baar Tasatonse Baar	Conter Co	LM LM LM LM LM LM LM LM LM LM	makes. The second seco	Mad () support shafting on U(L), or an exhand in U(T) Star () which, S(L), support shafting any protein basis and proteins. In advertise in any or allow 1.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2	1900 0607 0007 0007 0000 0000 0000 0000 0	2000         2000           2000         11704           2000         12704           2000         1270           2000         1200           20000	ear Market. I ear All Section of the Section of Section	Despiration Original OpenN OpenN OpenN Destination Destination Destination Produces AC OpenN Produces AC OpenN Produces AC OpenN AC OpenN AC OpenN Produces AC OpenN AC OpenN Produces AC OpenN
State         State           State         State           ST         State           ST2         State           State         State           State         State           State         State           State         State           State         State           State         State	GPT	Conter Co	LM LM LM LM LM LM LM LM LM LM	makes. Makes Makes Makes Makes Market States and Andres in a sing any any any any and phase summarizes the sing and any dependence in a second of the sing and any any any any and any	Mang Langung Handling and Kulu, Ju and Kunshandhi Ku Qiri Qiang Anthran CA, Serang Kang Mang Mang Mang Mang Tang Sang Anthran CA, Sang Mang Mang Mang Mang Mang Mang Mang Mang Mang Mang Mang Mang	(2000 0.0000 0.00000 0.00000 0.000000	2838           2838           2838           2838           2739           2739           2739           2739           2739           2739           2739           2739           2739           2739           2739           2739           2730	eer Maaks. eer Maaks. Her Maaks aan de personalen is en bier process Hangemoord for an open open open open open open open ope	Degetted Origined Origined Original Original Deathork Original Antionek Antionek Antionek
Saster Saster Saster St 271 273 273 273 273 273 273 273 273 273 273	007 Cogw 07 07 07 07 07 07 07 07 07 07	Conter Co	LM LM LM LM LM LM LM LM LM LM	makes ma	Mad () support shafting on U(L), or an exhand in O(T) Series y when C(A), we may any exhapting provide history and any exhapting of the strength provide history and any exhapting of the strength provide history and any exhapting of the strength of the strength of the strength of the strength of the strength of the strength of any exhapting of the strength of the strength of the strength and any exhapting of the strength of the strength of the strength of the strength of the strength of the strength of any exhapting of the strength of the strength of the strength of any exhapting of the strength of the strength of the strength of any exhapting of the strength of the strength of the strength of any exhapting of the strength of the strength of the Comparison strength or strength of the strength of the Comparison strength of the strength of the Comparison strength of the strength of the Comparison strength of the strength of the Strength of the strength of the strength of Comparison strength of the strength of Comparison strength of the strength of the Comparison strength of the strength of Strength of the strength of the strength of Strength of the strength of the strength of Strength of the strength of the strength of the Strength of the strength of the strength of the Strength of the strength of the strength of the Strength of the strength of the strength of the strength of Strength of the strength of the strength of the strength of Strength of the strength of the strength of the strength of the Strength of the strength of the strength of the strength of the Strength of the strength of the strength of the strength of the Strength of the strength of the	19900 06000 000000	2000         2000           2000         11734           1 500         11734           2 200         200           2 200         200           2 200         200           2 200         200           2 200         200           2 200         200           2 200         200           2 200         200           2 200         200           2 200         200           2 200         200           2 200         200           2 200         200           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320           2 200         21320	eer Meesk. Eer Meesk. Here Scher Lee see staffe Same Janvander 14 to 15 fe Janvan Hangemenn Frankrike en een staffe Same Janvander Frankrike en een staffe same Janvan Hangemenn Frankrike en een staffe Same Janvan Kannen Kannen Hangemenn Hangemenn Frankrike en een staffe Same Janvan Kannen Hangemenn Hangemenn Frankrike en een staffe Same Janvan Hangemenn Hange	Despiris Original OpenN OpenN OpenN Destroyk Des
State         State           State         State           Str         Str	007 009 007 007 007 007 007 007 007 007	Conter Conser Co	LM LM LM LM LM LM LM LM LM LM	makes. Makes and the second s	Mang Langung Handing and Kulu, Ju and Kunhan Mang Too Sang Antherin CAS, and an angen parameterina and "Sang Antherin CAS, and and an angen parameterina and "Sang Antherin Cash and and the sang parameterina and and an angen and angen and angen and angen and angen angen "Sang and angen angen and angen angen angen angen "Angen angen angen angen angen angen angen angen "Angen angen angen angen angen angen angen angen "Angen angen angen ang angen angen angen "Angen angen angen ang angen angen angen "Angen angen angen ang angen angen angen "Angen angen angen angen angen angen "Angen angen angen angen angen angen "Angen angen angen angen angen angen angen "Angen angen angen angen angen angen angen angen angen "Angen angen angen angen angen angen angen angen angen angen "Angen angen angen angen angen angen angen angen angen angen "Angen angen angen "Angen angen angen angen angen an	12000 0.0000 0.0000 0.00000 0.00000 0.000000	2880         2880           2880         1580           152         152           153         200           270         201           270         201           200         200           200 <td>ear Market. ear Market. Here In General and earlier and prevention in the In-Difference on the International Int</td> <td>Despired Origined Oceanit Ocea</td>	ear Market. ear Market. Here In General and earlier and prevention in the In-Difference on the International Int	Despired Origined Oceanit Ocea
Saster Saster Chi Off Off Off Off Off Off Off Off Saster S	007 Cogw 07 07 07 07 07 07 07 07 07 07	Conter Co	LM LM LM LM LM LM LM LM LM LM	makes. The second seco	Mad ju zapaga skaladi ga Kuju, za na oskaladi ku Off Song valanci, S.V., sang valanci ka zapaga prostek bara Sang prosteku ka zakaga ka zaka bara sang prosteku ka zakaga ka zaka bara sang valanci ka zakaga ka zaka bara bara bara sang valanci ka zakaga ka zaka bara bara bara sang valanci ka zakaga ka zaka bara bara bara sang valanci ka zaka bara bara bara bara bara bara sang valanci ka zaka bara bara bara bara bara bara sang valanci ka zaka bara bara bara bara Sang bara da sang valang sang sang sang sang sang sang sang s	19900 06000 06000 06000 06000 06000 070000 07000000	2830         2830           2830         1530           1 1731         1 531           2 153         1 53           1 53         1 53           2 10         1 53           2 10         1 53           2 10         1 53           2 10         1 53           2 201         1 53           2 202         1 53           2 203         1 53           2 204         1 53	eer Meesk. Eer Meesk. Here Scher Lee and Scher Lee and Here Annuelle Te do Lee and Here Meesk and Here Meesk Hangement Provide on Internet Scher Meesk and Here Meesk and Here Meesk Hangement Provide and Here Meesk and Here Meesk and Here Meesk Here Jack and Jack and Here Meesk and Here Meesk and Here Meesk Here Jack and Here Meesk and Here Meesk and Here Meesk and Here Meesk Here Jack and Here Meesk and Here Meesk and Here Meesk and Here Meesk Here Jack and Here Meesk and Here Meesk and Here Meesk and Here Meesk Here Jack and Here Meesk and Here Meesk and Here Meesk and Here Meesk Here Jack and Here Meesk and Here Meesk and Here Meesk and Here Meesk Here Jack and Here Meesk and Here Meesk and Here Meesk and Here Meesk Here Jack and Here Meesk and Here Meesk and Here Meesk and Here Meesk Here Jack and Here Meesk and Here Meesk and Here Meesk and Here Meesk and Here Meesk Here Jack and Here Meesk and Here Meesk and Here Meesk and Here Meesk Here Meesk and Here Meesk Here Meesk and Here Meesk Here Meesk and Here	Desprint Desprint Gentral
State         State           State         State           Str         Str	077 069 07 07 07 07 07 07 07 07 07 07 07 07 07	Conter Conser Co	LM LM LM LM LM LM LM LM LM LM	makes. Makes and the second s	Mang Langung Handling and Kulu, Ju an unknahls hu Off Gang subhan, GA, ganag Handling and Kulu. Juan Tang Juan Juan Juan Juan Juan Juan Juan Juan	42000 42007 4007 4	283         283           283         283           283         153           153         153           158         158           200         158           200         158           200         200	ear Market. ear Market. Here Scher Jack and Scher Scher Scher Jack and Scher	Dagettal Dagettal Casavi Casavi Casavi Dagetal Dagetal Dagetal Dagetal Dagetal Casavi Pantocal Casavi Pantocal Casavi Pantocal Casavi C
Saster SasterCa Ort Ort Ort Ort Sater Ort Sater Ort Sater Ort Sater Sater Ort Sater Ort Sater Ort Sater Ort Sater Ort Ort Sater Ort Ort Ort Ort Ort Ort Ort Ort Ort Or	047 047 047 047 047 047 047 047	Conter Co	Lui Lui Lui Lui Lui Lui Lui Lui Lui Lui	makes. Makes Market State of the second state and second state of performance is a second state of the second state of the second state of the second state of the second state of the second state of the second state of the second state of the second state of the second state state of the second state of the second state of the second state of the second state state of the second state of the second state of the second state of the second state of the second state state of the second state of th	Mad () support shalling on U(L), or an ownershift in O(T) Step systems, CA, Sector I service and a sector shall be of the Step systems, CA, Sector I service and Sector	12000 300000 300000 30000 30000 300000 30000 30000 30000 30000	203         203           203         203           1 101         101           1 201         101           1 201         101           1 201         101           1 201         101           2 201         101           2 201         101           2 201         101           2 201         201           2 202         202           2 203         203           2 203         203           2 204         203           2 204         203           2 204         204	eer Weeks. Yee The Grant of Least and Section of the Grant Branch of Least Annual Market Market Hangement Provide and Annual Handbook (1998). It is a section of Least Annual Market Market Hangement Provide Annual Market Market Market Market Market Market Market Market Market Market Market Market Market Market Market Market Market Phys. 14:053 doi:10.1007/11.000141. Phys. 14:050 doi:10.1007/11. Phys.	Despiration Despiration Quantum Quantum Despiration Despiration Despiration Produces Despirations ACT ACT ACT ACT ACT ACT ACT ACT ACT ACT
State         State           State         State           Str         Str	077 077 077 077 077 077 077 077 077 077	Conter Conser Co	Lui Lui Lui Lui Lui Lui Lui Lui Lui Lui	makes. Makes and the second s	Mad ja segarapit skulle get ULU, ja ze in schulet in Off Cogn partiers. Off, segarapit skulle get ULU, ja ze in schulet in Off Cogn partiers. At segarapit is knyre skullet VS Searapit en te Star skullet Searapit is knyre skullet VS Searapit en te Searapit skullet skullet in knyre skullet VS Searapit en te Searapit skullet skullet in knyre skullet VS Searapit en te Searapit skullet skullet in knyre skullet VS Searapit en te Searapit skullet skullet in knyre skullet VS Searapit en te Searapit skullet skullet skullet in skullet in Knyre skullet skullet skullet skullet in skullet in Knyre skullet skullet skullet skullet in Knyre skullet skullet skullet skullet skullet in Knyre skullet skullet skullet skullet skullet in Knyre skullet skullet skullet skullet skullet skullet Skullet schullet skullet skullet skullet skullet Skullet schulet skullet skullet skullet skullet Skullet skullet skullet skullet skullet skullet skullet skullet skullet skullet skullet Skullet skullet sk	0003 - 00	2 243 245 245 245 245 245 245 245 245	ear Market. Ear Market. Here Schoppen and	Dagettal Dagettal Dagettal Caratol Caratol Dagetal Dagetal Dagetal Dagetal Dagetal Dagetal Caratol Car
Saster Saster Chi Off Off Off Off Off Off Same Same Same Same Same Same Same Same	онт Содик Содик Сот Сот Сот Сот Сот Сот Сот Сот	Conter Co	Lui Lui Lui Lui Lui Lui Lui Lui Lui Lui	makes. Market and the second	Mad () support shading on U(L), or an ownerskin to O(T) Step system, CA, Sk and step any procedure take The periodic shading the step any procedure take the periodic shading the step any shading takes the step and step and step any shading the step any shading any shading the step any shading takes and any shading takes any shading the step any shading takes and any shading takes any shading the step any shading takes and any shading takes any shading the step and the shading takes and any shading takes any shading the step and take and any shading takes any shading the step and take and any shading takes any shading the step and take allows and shading takes any shading takes and any shading takes any	42000 42000 42000 42000 42000 42000 42000 42000 42000 42000 42000 42000 42000 42000 42000 42000 42000 42000 42000	243         243           11         243           12         243           13         243           14         243           15         243           15         243           15         243           16         243           17.0         243	eer Monet. I een Gapera ale selfel devel preventiel fa to UP proves I magement Proving on the Conference of the Section of t	Oragental Oppental Op
State         State	047 047 047 047 047 047 047 047	Conter Co	Lui Lui Lui Lui Lui Lui Lui Lui Lui Lui	makes. The second seco	Mang Langung Handling and Kulu, Ju and Kunhandhi Kuli QYT Chang Anthron, G.M., and Kuli Langung Handhin Kulu Tang Jang Kuli Langung Handhing Kulu Tang Jang Kuli Langung Handhing Kulu Tang Jang Kuli Langung Handhing Kuli Langung Handhing Tang Jang Langung Handhing Kuli Langung Handhing Kuli Langung Handhing Kuli Handhing Kuli Langung Handhing Cangang Langung Handhing Langung Handhing Cangang Langung Handhing Kuli Handhing Kuli Handhing Kuli Langung Handhing Kuli Handhing Kuli Handhing Cangang Jang Handhing Kuli Handhing Kuli Handhing Cangang Langung Handhing Kuli Handhing Kuli Handhing Kuli Handhing Kuli Handhing Kuli Handhing Kuli Handhing Kuli Handhing Sanama Giff Sanama Gali Kuli Handhing Kuli Handhing Sanama Gali Kuli Handhing Kuli Handhing Kuli Handhing Sanama Gali Kuli Handhing Kuli Handhing Kuli Handhing Kuli Handhing Kuli Handhing Kuli Handhing Sanama Gali Kuli Handhing Kuli	4000 4000 4000 4000 4000 4000 4000 400	243         243           243         243           1         25           1         15           1         15           1         15           1         15           1         15           1         15           1         15           1         15           2         15           2         15           2         15           2         15           2         15           2         15           2         15           2         15           2         15           2         15           2         15           2         15           3         15           3         15           3         15           3         15           3         15           3         16           3         16           3         16           3         16           3         16           3         16           3         16	ear Market. Ear Market Section 2014 Section 2014 Section 2014 Section 2014 International Control 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2015 Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Section 2014 Section 2014 Section 2014 Section 2014 International Section 2014 Se	Despiration Despiration Description Descr
		Conter Co	ud ud ud ud ud ud ud ud ud ud ud ud ud u	makes. Market and the second	Mang Language strately on ULU, to an ownerskin to OTT Desperation, S. Language strate that a strategy any protect has a strategy may protect has a strategy and the strategy any protect has a strategy any prot	1000 03000 0000 0000 0000 0000 0000 000	283         283           1         293           1         19           2         20           1         19           2         20           1         19           2         20           1         19           2         19           2         20	ier model. Ier mo	Desprint Operation Operati
Saster Saster Grit 271 273 273 273 273 273 273 273 273 273 273	047 047 047 047 047 047 047 047	Conter Co	Lui Lui Lui Lui Lui Lui Lui Lui Lui Lui	makes. Market and the second	Mad (a sequence) shading out 100, ju an outwards the OPT Consequences of the sequence shares the outward outwa	1000 03000 04000 03000 04000000	2 233 243 243 243 243 253 253 253 253 253 253 253 25	eer measter. Learn Gaper, Learn and Learner and Learn	Despiration Constration Constration Constration Enclosed Enclosed Constration
Saster Saster Grit Grit Grit Grit Grit Grit Grit Gri		Conter Co	UI U	makes. Makes Makes Market States and Andres makes to go and physics are states of physics and physics Market States and Andres makes and physics and physics and physics and physics Market States and Andres and Physics and Physics and Physics and Physics Market States and Physics and Physics and Physics and Physics and Physics Market States and Physics and Physics and Physics and Physics and Physics Market States and Physics and Physics and Physics and Physics and Physics Market States and Physics and Physics and Physics and Physics and Physics Market States and Physics and Physics and Physics and Physics and Physics Market States and Physics and Physics and Physics and Physics and Physics Market States and Physics and Physics and Physics and Physics and Physics Market States and Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and Physics and Physics and Physics and Physics Market Physics and	Mang Langung Handhing and Yuli, Ju and Kanaharahan Orti Yang Yanhimo CA, Bernard Hannary, Bernard Hann, Yan Hannarahan, Yan Langung Handhing Hannarahan, Yan Hannarahan, Yangang Handihan, Yan Hannarahan, Yan		243 245 245 245 245 245 245 245 245	iner Wande. I en Stepen Zaus et al. Same a provinci in the Sur Province I hangement of the same a specific stress of the same and the same and the same and international stress of the same and the same and the same and the same and in the same and the same and the same and the same and the same and the same and the same and the same and the same and the same and the same and the same and the same and the same and the same and the same and the same and the same and the sa	Despiration Consention Consention Consention Enclander Enclander Consention C
State         State	047 04940 047 047 047 047 047 047 047 0	Conter Co	ud au au au au au au au au au au	makes. The second seco	Mang Langung Hashing and Kulu, Ja and Kunhan Ku (1977) Sang Sahahar, Sala, Sang Sahahar, Sang Sang Sahahar, Sang Sang Sahahar, Sang Sang Sang Sang Sang Sang Sang Sang	00000 00000 00000 00000 00000 00000 0000	243         243           743         743           1 (1)         1 (1)           1 (2)         210           1 (2)         210           1 (2)         210           2 (2) <td>ear Market. Team Schart Law Set of Same Japanetic 1 is to UP annue in the south table in the south tables of the south tables</td> <td>Desprint Cogenital Cogenital Cogenital Cogenital Enutoretic Cogenital Cogeni</td>	ear Market. Team Schart Law Set of Same Japanetic 1 is to UP annue in the south table in the south tables of the south tables	Desprint Cogenital Cogenital Cogenital Cogenital Enutoretic Cogenital Cogeni
State         State	онт орту орту орту орт орт орт орт орт орт орт орт	Inder Construction of the second of the seco	Lui du lu	makes. The second seco	Negl spages sharing and table per table per sense methods the OPT Operation of the Sense sense sense share and operation of the Sense sense sense and the Sense sense sense sense sense sense sense sense sense response sense share sense the sense sense sense sense sense response sense sense sense sense sense sense sense sense sense response sense sense sense response sense sense sense sense sense sense sense sense response sense sense sense sense sense sense sense sense sense response sense sense sense sense sense sense sense sense sense response response response sense separation response sense sense sense response response response response response sequences response response response resp	0000 0000 0000 0000 0000 0000 0000 0000 0000	2 283 295 295 295 295 295 295 295 295	ear Market. Ear Market. Here School and an earlier and provided in the School Provide and Market Andread School Here School and School and School Andread School Andread School Andread School Andread School Here School Andread Andread School Andread School Andread School Andread School Andread Andread Andread School Andread Andrea	Desprint Cogenital Cogenital Cogenital Cogenital Enclosed Cogenital Cogenita
Sater Sater		Seater Se	U           U	makes. Makes and the second s	Mang Langung Handling and Kulu, Ju and Kunhandhi Kuli QYT Carl synthesis CAL, San Kuli Hang Langung Handhi Kuli Kuli Tang Langung Langung Langung Langung Langung Langung Tang Langung Langung Langung Langung Tang Langung Langung Langung Langung Langung Langung Langung Langung Langung Tang Langu	00000000000000000000000000000000000000	243         243           783         783           1 (1)         1 (1)           1 (2)         1 (2)           1 (2)         20 (1)           1 (2)         20 (1)           2 (2)         20 (1)           2 (2)         20 (1)           2 (2)         20 (1)           2 (2)         20 (1)           2 (2)         20 (1)           2 (2)         20 (1)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)           2 (2)         20 (2)	ear Market. Team Schart Law Section S	Desprint Operation Operati
Saster Saster Saster Sart Sart Sart Sart Sart Sart Sart Sar	онт орти орти орти орт орт орт орт орт орт орт орт	Senter Carlor of	ui du lui	makes. Makes M	Negl spages sharing and table per table per sharing and per spage sharing and per spage sp	0000 0000 0000 0000 0000 0000 0000 0000 0000	243         243           243         243           1         24           1         23           1         23           1         23           1         24           1         25           1         26           1         26           1         26           2         26	ier Market. Ier M	Dagnid Capanid Capanid Capanid Capanid Examinet Capanid Capani
Saster Saster Saster Sat 271 273 273 273 273 273 273 273 273 273 273		Sender Desaf Color	ui di	makes. Makes and the second s	Mad (a support shafting on U.G.), and an exhand in OPT Series yorks (M.G. (S. Arrowski), and an exhange support the interpretation of the support of		243         243           243         243           243         243           243         243           1         24           1         24           243         243           243	eer Theode. Term Conjunct and Conference and Confe	Despiration Cognit
Seater Seater Seater Ser 2017 2013 2017 2017 2017 2017 2017 2017 2017 2017	онт онт онт онт онт онт онт онт	Conter Conse	Lui du lu	makes. The second seco	Mang Langung Handing and Kulu, Ja an an and walk in QPT Que yatherio, GA, Kanan Y,		243         243           243         243           1         24           1         23           1         23           1         23           1         23           1         23           1         24	ear Market. Ear Market. Ear Market Section of Section	Desprint Control Contr
Saster Saster Saster ST ST ST ST ST ST ST ST ST ST ST ST ST		Conter Co	ui ou	makes. Makes and the second s	Mang Langung Handling and Kali, Ju and Kanhal Hu. Off. 2014 yearshine, S.G. Samara, S.M. Shang and Kali. Samara The anti-section of the section of the sec		243         243           244         243           243         243           243         243           243         243           243	ear Market. Team Schurz Las selfel Sama Jarveski fra Sul Jarveski Frankrik Hangemen Kaller Sama Selfer Sama Jarveski fra Sul Jarveski Frankrik Hangemen Kaller Sama Selfer Sama Jarveski frankrik Hangemen Kaller Sama Selfer Sama Selfer Sama Selfer Sama Selfer Hangemen Kaller Sama Selfer Sama Selfer Sama Selfer Sama Selfer Hangemen Kaller Sama Selfer Sama Selfer Sama Selfer Sama Selfer Hangemen Kaller Sama Selfer Sama Selfer Sama Selfer Sama Selfer Sama Selfer Hangemen Kaller Sama Selfer Sama Selfer Sama Selfer Sama Selfer Hangemen Kaller Sama Selfer Sama Selfer Sama Selfer Sama Selfer Sama Selfer Hangemen Kaller Sama Selfer Sama Selfer Sama Selfer Sama Selfer Sama Selfer Sama Selfer Hangemen Kaller Sama Selfer Sama Self	Desprint Operation Operation Operation Operation Description Control C
Seater Seater Seater Set 271 273 273 273 273 273 273 273 273 273 273	orr Goryan Goryan Gory Gr Gr Gr Gr Gr Car Car Car Car Car Car Car Car Car Ca	Conter Co	Jui	minute. Market and the second	Mang Langang Hashing and Kulu, Ja and Kunhan Bin Off. Series pathetics, Silv, and Kulu, Ja and Kulu, Shang and Kulu. The series of the serie		243         243           2         244           2         244           2         244           2         244           2         244           2         244           2         244           2	ier Wender. Ier Wender. Ier Konne in der pris seit auf ihr and personalis i hen folge prises Insegnenze Wender i hen der Sterner Sterner i Henner Henner Henner Henner Insegnenze Wender i henner Sterner Sterner i Henner	Desprint Control Control Control Control Evaluation Evaluation Control
Saster Saster Chi Saster Saste	orr Orr Orr Orr Orr Orr Orr Orr	Inster Control of Cont	ui ou	minutes. The second se	Negl spega thanking and table part and a monarchick to OPT Organ spectra (SA). See a spectra of the spectra of		200         200           200	iner Market.  In Concern Section 2014 In the Concern Constant Section 2014 In the Concern Constant Section 2014 In the Constant 2014 In the Constant Section 2014 In the Constant Section 2014 In the Constant Section 2014 In the Constant 2014 In the Con	Despiration Despiration Description Descri
State         State		Instart Control of Con	UI	minute. minute in the second	Negl spages planking on ULU, a non-sharkin DYT Serg valetta, OL, a spage planking on Serg and Serg Serg Table Serg valetta, OL, a spage space has no spage and space space and space space space space space space space space space space space space space space space space space space space Table space space space space space space space space space space Table space		243         243           253         243           254         253           254         254           254         254           254         254           254         254           254         254           254         254           254         254           254         254           254         254           254         254           254	ear Market. Team Scher Les auf Les and provide in the Suff Provide provide Network of Suff Scher Suff Provide provide in the Suff Provide provide Network of Suff Scher Suff Provide Provide Scher Suff Provide Scher Suff Scher Suff Provide Scher Suff Scher Suff Provide Scher Suff Provide Scher Suff Scher Suff Scher Suff Scher Suff Scher Suff Provide Scher Suff Sch	Dagning J Capaning J C
Saster Saster Chi Office Safet	orr orr orr orr orr orr orr orr orr orr	Instart Dis	Jui	minute. Market and the second	Nach Langung Hanking and Kulu, Ja and Kunka Hu (1977) Sang sathers CAS, Sang Sang Sang Sang Sang Sang Sang Sang		243         243           244         243           243         243           243         244           244         244           244         244           244         244           244         244           244         244           244         244           244         244           244         244           244	ier Warden. See The Galaxy and an effect of the Section of the Section of Texastrate moves that equals the section of the Section of Section o	Dagmid Capacital
Saster Sa	orr Orgense Orr Orr Orr Orr Orr Orr Orr Corr Corr C	Contert	За	minutes in the second s	Mang Langung Hashing and Kulu, Ju and Kunhan Mang Ort Toga spatience, Odk, and an interpret provide history and an interpret provide history of the XLB hash produces for a spatience hash and an interpret production and and a spatience hash and and the spatience hash and a spatience hash and and the spatience hash and an and the spatience hash and a spatience hash and an and the spatience hash and a spatience hash and and a spatience hash		200         200           200	ier Warder. Ier Warder. Hen Gogen zur seint framen personiel ihr is Giff anna en term in the south Warder. Hen Gogen zur seint framen personiel ihr is Giff anna en term in the south in the south Henderscher Strategie and Strategie and Strategies and Personiel and and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Strategies and Strategies and Strategies and Strategies and Strategies and Henderscher Strategies and Stra	Dagnidi Capanidi Capanidi Capanidi Capanidi Enutomik Capani Capani Capani Anti Capani Anti Capani Anti Capani Anti Capani Anti Capani Anti Capani Cap
Saster SesterCan Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser	orr organization organization organization orr orr orr orr orr orr orr orr orr o	Conter       Conter <td< td=""><td>Jui</td><td>makes. Makes Makasa Makasa Makes Makes Makes Makes Makes Makes Makes</td><td>Na () segue share () a set (u), i a no share is no (r)? Segue share, () a segue share is no share is no (r)? Segue share, () a segue share is no (r). I a segue share is no segue share () a segue share is no (r) a segue share is no segue share () a segue share is no (r) a segue share is no segue share () a segue share () a segue share is no (r) a segue () a segue share () a segue share is no (r) a segue share is no segue share () a segue share () a segue share is no segue share () a segue share () a segu</td><td></td><td>200         200           200         200           170</td><td>ier Wenke. See The Galaxy and set of annu provided in the Galaxy formation much than the Galaxy and The Galaxy and Galaxy and Calaxy and The Galaxy formation much that galaxy formation the Galaxy and Galaxy and Galaxy (Calaxy Calaxy) and Calaxy and Calaxy (Calaxy Calaxy) The Le Galaxy and Calaxy (Calaxy Calaxy) and Calaxy (Calaxy Calaxy) The Le Galaxy and Calaxy (Calaxy Calaxy) and Calaxy (Calaxy Calaxy) The Le Galaxy and Calaxy (Calaxy Calaxy) and Calaxy (Calaxy Calaxy) The Le Galaxy and Calaxy (Calaxy Calaxy) The Calaxy (Calaxy Cala</td><td>Dagnid Ca</td></td<>	Jui	makes. Makes Makasa Makasa Makes Makes Makes Makes Makes Makes Makes	Na () segue share () a set (u), i a no share is no (r)? Segue share, () a segue share is no share is no (r)? Segue share, () a segue share is no (r). I a segue share is no segue share () a segue share is no (r) a segue share is no segue share () a segue share is no (r) a segue share is no segue share () a segue share () a segue share is no (r) a segue () a segue share () a segue share is no (r) a segue share is no segue share () a segue share () a segue share is no segue share () a segue share () a segu		200         200           200         200           170	ier Wenke. See The Galaxy and set of annu provided in the Galaxy formation much than the Galaxy and The Galaxy and Galaxy and Calaxy and The Galaxy formation much that galaxy formation the Galaxy and Galaxy and Galaxy (Calaxy Calaxy) and Calaxy and Calaxy (Calaxy Calaxy) The Le Galaxy and Calaxy (Calaxy Calaxy) and Calaxy (Calaxy Calaxy) The Le Galaxy and Calaxy (Calaxy Calaxy) and Calaxy (Calaxy Calaxy) The Le Galaxy and Calaxy (Calaxy Calaxy) and Calaxy (Calaxy Calaxy) The Le Galaxy and Calaxy (Calaxy Calaxy) The Calaxy (Calaxy Cala	Dagnid Ca
Sater Sater	orr Orgen Orr Orr Orr Orr Orr Orr Orr Orr Auto Orr Tautores Auto Corr Auto Corr Auto Auto Corr Corr Auto Corr Corr Corr Corr Corr Corr Corr Co	Seater       Design       Obser	Ju J	minute. minute in the second	Negl spages sharing out ULL to an extraction UPT Design space. OL A space space space space space The space space space space space space space space space The space space space space space space space space space The space space space space space space space space space space The space spac		200         200           200	ior Water. The Source of the	Dagnisi Cagarist Cagarist Cagarist Cagarist Cagari
Sater Settor Set	orr orr orr orr orr orr orr orr orr orr	Context	Jui - Sui -	makes. Makes M	Na () segue share () a set (u), i a no share is no (r)? Segue share, () a segue share is no share is no (r)? Segue share, () a segue share is no (r). I a segue share is no segue share () a segue share is no (r) a segue share is no segue share () a segue share is no (r) a segue share is no segue share () a segue share is no (r) a segue share is segue share () a segue share is no (r) a segue share is Segue share () a segue share is no (r) a segue share is Segue share () a segue share is no (r) a segue share Segue share () a segue share () a segue share Segue share () a segue share is no (r) a segue share Segue share () a segue share () a segue share Segue share () a segue share () a segue share Segue share () a segue share () a segue share Segue share () a segue share () a segue share Segue share () a segue share () a segue share () a segue share Segue share () a segue share () a segue share () a segue share Segue share () a segue sha		200         200           200         200           170         200           170         200           200	ier Wender. Ier Wender. Ier Konner Seiner	Dagnid Cagnid Cardi Card
Sater Sater	orr Orgener Orr Orr Orr Orr Orr Orr Orr Orr Tautores Autor A	Seater       Seater <td< td=""><td>Ju Ju J</td><td>minute minute</td><td>Negl spages sharing out Nui, Ju an exhanish mo (PT) Series spaces. Out Nui and Series and Series Ser</td><td></td><td>200         200           200</td><td>ier Washe. See Subject See Subject Su</td><td>Desprint Cogenital Cogenit</td></td<>	Ju J	minute minute	Negl spages sharing out Nui, Ju an exhanish mo (PT) Series spaces. Out Nui and Series and Series Ser		200         200           200	ier Washe. See Subject See Subject Su	Desprint Cogenital Cogenit
Saster SesterCas ort ort ort ort orts orts orts orts or	orr orr orr orr orr orr orr orr orr orr	Content	Jui du la	makes. Makes M	Na () segura plan blank gen t U.). Ja an unkan kin (1977 Segura plants). Ok (a segura plant kin kin (1978) Segura plants). Ok (a segura plant kin kin (1978) Segura plants) and kan (1978) segura plant kin (1978) Segura plants) segura plant kin (1978) segura plant kin (1978) Segura plants) segura plant kin (1978) segura plant kin (1978) Segura plants) segura plants) Segura plants) segura plants (1978) Segura plants) segura plants) Segura plants) segura plants (1978) Segura plants) segura plants) segura plants) Segura		200         200           200         200           170         200           120         200           200	ier Market. Ier M	Dagniti Capa
Sate Characteries Sate Charact	orr Org Orr Orr Orr Orr Orr Orr Orr Orr Corr Table Orr Table Orr Table Orr Autor Orr Autor Orr Autor Orr Autor Orr Autor Orr Orr Autor Orr Orr Autor Orr Orr Autor Orr Orr Orr Autor Orr Orr Orr Orr Orr Orr Orr Orr Orr O	Searce       Searce <td< td=""><td>Ju Ju J</td><td>minutes in the second s</td><td>Negl spages sharing out to be a set out and the OPT Conservation of the Second Second</td><td></td><td>200         200           200</td><td>ear Market. The Source of the Source of the</td><td>Desprint Cogenital Cogenit</td></td<>	Ju J	minutes in the second s	Negl spages sharing out to be a set out and the OPT Conservation of the Second		200         200           200	ear Market. The Source of the	Desprint Cogenital Cogenit

Figure 5: You can access the original table at https://docs.google.com/spreadsheets/d/ 1ltyrAB6BL29cOv2fSpNQnnq2vbX8UrH147d7FkIf6t4 for easier browsing across the different model features.



Figure 6:



Figure 7: Transformer timeline. Colors describe Transformer family.



Figure 8: Transformer timeline. On the vertical axis, number of parameters. Colors describe Transformer family.

- Family: SE(3) Transformer<sup>21</sup>
- Pretraining Architecture: Encoder
- **Pretraining Task:** Protein folding prediction\*ion of BERT using parameter sharing, which is much more efficient given the same number of parameters
- Extension: The original Alphafold used a BERT-style transformer. The details of Alphafold's Transformer are not known, but it is believed it is an extension of the SE(3)-Tranformer, a 3-D equivariant Transformer (see this blog post<sup>22</sup>)
- Application: Same as BERT
- Date (of first known publication): 09/2019
- Num. Params:b12M, Large = 18M, XLarge = 60M\*
- Corpus:Same as BERT
- Lab:Google
- 2.5.3 Anthropic Assistant
  - **Reference:**<sup>23</sup> see also<sup>24</sup> [13, 14]
  - Family: GPT
  - Pretraining Architecture: Decoder
  - **Pretraining Task:** Protein folding prediction\*ion of BERT using parameter sharing, which is much more efficient given the same number of parameters
  - Extension: These models do not introduce novelties at the architecture/pretraining level and they are based on GPT-3 but rather focuses on how to improve alignment through fine-tuning and prompting. Note that the Anthropic Assistant includes several models optimized for different tasks. Latest versions of this work focus on the benefits of RLHF

<sup>&</sup>lt;sup>21</sup>https://arxiv.org/abs/2006.10503

<sup>&</sup>lt;sup>22</sup>https://fabianfuchsml.github.io/alphafold2/

<sup>&</sup>lt;sup>23</sup>https://arxiv.org/abs/2112.00861

<sup>&</sup>lt;sup>24</sup>https://arxiv.org/abs/2204.05862

- Application: Different models with different applications from general dialog to code assistant.
- Date (of first known publication): 12/2021
- Num. Params:10M to 52B
- **Corpus:**400B tokens from filtered Common Crawl and Books. They also create several Dialogue Preference datasets for the RLHF training.
- Lab:Anthropic

# 2.5.4 BART

- **Reference:**<sup>25</sup>[15]
- Family: BERT for encoder, GPT for Decoder
- Pretraining Architecture: Encoder/Decoder
- Pretraining Task: DAE
- Extension: It can be seen as a generalization of BERT and GPT in that it combines ideas from both in the encoder and decoder
- Application: Mostly text generation but also some text understanding tasks\*
- Date (of first known publication): 10/2019\*
- Num. Params:10 % more than BERT
- Corpus:Same as RoBERTa (160Gb of news, books, stories
- Lab:Facebook

# 2.5.5 BERT

- **Reference:**<sup>26</sup>[16]
- Family: BERT
- Pretraining Architecture: Encoder
- Pretraining Task: MLM/NSP
- Extension: It can be seen as a generalization of BERT and GPT in that it combines ideas from both in the encoder and decoder
- Application:General Language Understanding and Question Answering. Many other language applications followed
- Date (of first known publication): 10/2018
- Num. Params:Base = 110M, Large = 340MT
- Corpus:Toronto Book Corpus and Wikipedia (3.3B Tokens)
- Lab:Google

# 2.5.6 Big Bird

- **Reference:**<sup>27</sup>[17]
- Family: BERT
- Pretraining Architecture: Encoder
- Pretraining Task: MLM
- Extension: Big Bird can extend other architectures such as BERT, Pegasus, or RoBERTa by using a sparse attention mechanism that elminates the quadratic dependency thus making it more suitable for longer sequences
- Application: Particularly well suited for longer sequences, not only in text but also e.g. in genomics
- Date (of first known publication): 07/2020

<sup>&</sup>lt;sup>25</sup>https://huggingface.co/docs/transformers/model\_doc/bart

<sup>&</sup>lt;sup>26</sup>https://huggingface.co/docs/transformers/model\_doc/bert

<sup>&</sup>lt;sup>27</sup>https://huggingface.co/docs/transformers/model\_doc/big\_bird

- Num. Params: Depends on the overall architecture
- Corpus:Books, CC-News, Stories and Wikipedia)
- Lab:Google

#### 2.5.7 BlenderBot3

- **Reference:**<sup>28</sup>[18]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: BlenderBot 3 is based on a pre-trained OPT. It adds features needed for a dialog agent such as long-term memory or the ability to search the internet. It is also fine-tuned for some specific tasks given human feedback on them.
- Application: Same as GPT-3
- Date (of first known publication): 08/2022
- Num. Params:175B
- Corpus: 180B tokens = RoBERTa + the Pile + PushShift.io Reddit
- Lab:Facebook

### 2.5.8 BLOOM

- **Reference:**<sup>29</sup>
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Main difference to GPT-3 is that it uses full attention instead of sparse attention
- Application: Same as GPT-3
- Date (of first known publication): 07/2022
- Num. Params:176B
- Corpus: 366B tokens (1.5 TB of text data) multilingual dataset
- Lab: Big Science/Huggingface

# 2.5.9 ChatGPT

- Reference:<sup>30</sup>
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: ChatGPT takes a GPT3.5 (aka GPT3 Davinci-003) pretrained model and uses RLHF to finetune the model mostly like described in InstructGPT but with slight differences in the data collection. ChatGPT is also more than a model since it includes extensions for Memory Store and retrieval similar to BlenderBot3
- Application: Dialog agents
- Date (of first known publication): 10/2022
- Num. Params:Same as GPT3
- Corpus: Same as GPT3 + datasets generated for RLHF
- Lab: OpenAI

<sup>&</sup>lt;sup>28</sup>https://arxiv.org/abs/2208.03188

<sup>&</sup>lt;sup>29</sup>https://huggingface.co/docs/transformers/model\_doc/bloom

<sup>&</sup>lt;sup>30</sup>https://openai.com/blog/chatgpt/

### 2.5.10 Chinchilla

- **Reference:**<sup>31</sup>[19]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Same as Gopher but with optimizations to reduce model size and therefore training/inference time with equal or superior performance
- Application: Same as Gopher/GPT3
- Date (of first known publication): 03/2022
- Num. Params:70B
- Corpus: Massive Text
- Lab: Deepmind

# 2.5.11 CLIP

- Reference:<sup>32</sup>[20]
- Family: CLIP (Also using Resnet, ViT, and vanilla transformer for text)
- Pretraining Architecture: Encoder
- Pretraining Task: predict which of the N × N possible (image, text) pairings across a batch actually occurred
- Extension: Combines Resnet and ViT for the visual encoding with Transformer for the Textual encoder
- Application: Image/object classification
- Date (of first known publication): 02/2021
- Num. Params:?
- Corpus: WIT (WebImageText) 400 million text, image pairs
- Lab: OpenAI

# 2.5.12 CM3

- Reference:<sup>33</sup>[21]
- Family: HTML
- Pretraining Architecture: Decoder
- Pretraining Task: Causality-masked LM
- Extension: This is somewhat similar to HTML in its use of structured training data. However, it is a different architecture and uses causal masking
- Application: Multimodal language model with the ability to do structured prompting
- Date (of first known publication): 01/2022
- Num. Params:13B (largest)
- Corpus: CC-News, English Wikipedia
- Lab: Facebook

<sup>&</sup>lt;sup>31</sup>https://arxiv.org/abs/2203.15556

<sup>&</sup>lt;sup>32</sup>https://huggingface.co/docs/transformers/model\_doc/clip

<sup>&</sup>lt;sup>33</sup>https://arxiv.org/abs/2201.07520

# 2.5.13 CTRL

- **Reference:**<sup>34</sup>[22]
- Family:
- Pretraining Architecture: Decoder
- Pretraining Task:
- Extension: model can generate text conditioned on control codes that specify domain, style, topics, dates, entities, relationships between entities, plot points, and task-related behavior
- Application: Controllable text generation
- Date (of first known publication): xx
- Num. Params:1.63B
- **Corpus:** 140 GB of text including: Wikipedia (En, De, Es, Fr), Project Gutenberg, 45 subreddits, OpenWeb-Text2, Amazon Reviews, Europarl and UN data from WMT, question-answer pairs from ELI5, and the MRQA shared task3, which includes the Stanford Question Answering Dataset, NewsQA, TriviaQA, SearchQA, HotpotQA, and Natural Questions
- Lab: Salesforce

# 2.5.14 DALL-E

- Reference:<sup>35</sup>[23]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: Caption prediction
- Extension: A differential variational auto-encoder is used to learn the visual codebook. The transformer is a variation of GPT-3
- Application: Text to image
- Date (of first known publication): 01/2021
- Num. Params:12B
- Corpus: 250 million text-images pairs from the internet
- Lab: OpenAI

# 2.5.15 DALL-E 2

- **Reference:**<sup>36</sup>[24]
- Family: CLIP, GLIDE
- **Pretraining Architecture:** Encoder/Decoder
- Pretraining Task: Caption prediction
- Extension: Combines CLIP encoder and Diffusion decoder similar to GLIDE
- Application: Text to image
- Date (of first known publication): 04/2022
- Num. Params:3.5B
- Corpus: Combination of the DALL-E and CLIP datasets
- Lab: OpenAI

<sup>35</sup>https://openai.com/blog/dall-e/

<sup>&</sup>lt;sup>34</sup>https://huggingface.co/docs/transformers/model\_doc/ctrl

<sup>&</sup>lt;sup>36</sup>https://openai.com/dall-e-2/

### 2.5.16 Decision Transformers

- **Reference:**<sup>37</sup>[25]
- Family: GPT, Control Transformers" (not per se a family, but grouping here those transformers that try to model more general control, RL-like, tasks)
- Pretraining Architecture: Decoder
- **Pretraining Task:** Next action prediction
- Extension: Decision transformers use a GPT architecture and extend it by encoding trajectories in a way that they can be learned by an auto-regressive task
- Application: General RL (reinforcement learning tasks)
- Date (of first known publication): 06/2021
- Num. Params:Same as GPT
- Corpus: Different corpus for different experiments
- Lab: Google/UC Berkeley/Facebook

# 2.5.17 DialoGPT

- Reference:<sup>38</sup>[26]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: GPT-2 architecture trained on dialog data
- Application: Text generation in dialog settings
- Date (of first known publication): 10/2019
- Num. Params:1.5B
- Corpus: 140M Reddit conversations
- Lab: Microsoft

#### 2.5.18 DistilBERT

- **Reference:**<sup>39</sup>[27]
- Family: BERT
- Pretraining Architecture: Encoder
- Pretraining Task: MLM/NSP
- Extension: Compressed version of BERT using distillation, which is much more efficient given the same number of parameters
- Application: Same as BERT
- Date (of first known publication): 10/2019
- Num. Params:66M
- Corpus: Same as BERT
- Lab: Huggingface

<sup>&</sup>lt;sup>37</sup>https://arxiv.org/abs/2106.01345

<sup>&</sup>lt;sup>38</sup>https://huggingface.co/docs/transformers/model\_doc/dialogpt

<sup>&</sup>lt;sup>39</sup>https://huggingface.co/docs/transformers/model\_doc/distilbert

# 2.5.19 DQ-BART

- **Reference:**<sup>40</sup>[28]
- Family: BART
- Pretraining Architecture: Encoder/Decoder
- Pretraining Task: DAE
- Extension: Adds quantization and distillation to a BART model to improve performance and model size
- Application: Text generation and understanding
- Date (of first known publication): 03/2022
- Num. Params: Up to 30x reduction in parameters compared to standard BART
- Corpus: CNN/DM, XSUM, ELI5, WMT16 En-Ro (1M tokens)
- Lab: Amazon

# 2.5.20 ELECTRA

- **Reference:**<sup>41</sup>[29]
- Family:
- Pretraining Architecture: Encoder
- Pretraining Task: RTD
- Extension: Same as BERT
- Application: 03/2020
- Date (of first known publication): xx
- Num. Params:Base = 110M, Large = 330M
- Corpus: Same as BERT except for Large with is same as XLNet
- Lab: Stanford/Google

# 2.5.21 ERNIE

- **Reference:**<sup>42</sup>[30]
- Family: BERT
- Pretraining Architecture: Encoder
- Pretraining Task: MLM
- Extension: Uses BERT for Encoder architecture, but stacks and aggregates two of them for text and entities. This architecture could be understood as BERT for text + knowledge graphs
- Application: Knowledge intensive related tasks that might benefit from knowledge graphs or entities such as entity recognition
- Date (of first known publication): 05/2019
- Num. Params:114M
- **Corpus:** English Wikipedia + Wikidata for entitites (note that they initialize model to original BERT parameter values
- Lab: Various Chinese institutions

<sup>&</sup>lt;sup>40</sup>https://arxiv.org/abs/2203.11239

<sup>&</sup>lt;sup>41</sup>https://huggingface.co/docs/transformers/model\_doc/electra

<sup>&</sup>lt;sup>42</sup>https://arxiv.org/abs/1905.07129

# 2.5.22 Flamingo

- **Reference:**<sup>43</sup>[31]
- Family: Chinchilla
- Pretraining Architecture: Decoder
- Pretraining Task: Log likelihood of text given some visual input
- Extension: It uses a frozen textual language model (like Chinchilla) conditioned on the visual representation, which is encoded from a Normalizer-Free ResNet
- Application: Text to image
- Date (of first known publication): 04/2022
- Num. Params:80B (largest)
- **Corpus:** MultiModal MassiveWeb (M3W): 185 million images and 182 GB text + a number of text paired with image datasets: ALIGN + LTIP (Long Text & Image Pairs) = 312 million images, and VTP (Video & Text Pairs) = 27 million short videos (approximately 22 seconds on average)
- Lab: Deepmind

# 2.5.23 Gato

- **Reference:**<sup>44</sup>[32]
- **Family:** "Control Transformers" (not per se a family, but grouping here those transformers that try to model more general control, RL-like, tasks)
- Pretraining Architecture: Decoder
- **Pretraining Task:** MLM (where tokens are either text or agent actions)
- Extension: The standard decoder-only transformer architecture is preceded by an embedding layer that can embed text and images, plus add position encodings to add spatial information when applicable.
- **Application:** Gato presents a generalizable agent that can be used beyond text to tasks such as playing Atari or controlling a robot arm.
- Date (of first known publication): 05/2022
- Num. Params:1.2B
- **Corpus:** 1.5T tokens including standard text (e.g. MassiveText), vision (e.g. ALIGN), and simulation environments (e.g. ALE Atari, or RGB Stacking Real Robot)
- Lab: Deepmind

# 2.5.24 GLaM

- **Reference:**<sup>45</sup>[33]
- Family: Transformer
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: GLaM introduces a Mixture of 64 Experts to increase parameter count and generalization properties in a somewhat standard decoder-only. Transformer architecture. Only two experts get activated at a time per token, which makes the model also more efficient in training and inference.
- Application: General language modeling
- Date (of first known publication): 12/2021
- Num. Params: 1.2T across 64 experts, but only 96B get activated for inference
- Corpus: 1.6T tokens including web pages filtered by Wikipedia and books for quality
- Lab: Google

<sup>&</sup>lt;sup>43</sup>https://arxiv.org/abs/2204.14198

<sup>&</sup>lt;sup>44</sup>https://www.deepmind.com/publications/a-generalist-agent

<sup>&</sup>lt;sup>45</sup>https://ai.googleblog.com/2021/12/more-efficient-in-context-learning-with.html

# 2.5.25 GLIDE

- **Reference:**<sup>46</sup>[34]
- Family: Diffusion models
- **Pretraining Architecture:** Encoder
- Pretraining Task: Caption prediction
- Extension: GLIDE can be seen as an extension of the ADM (Ablated Diffusion Model) by the same authors. However, ADM is not per se a transformer architecture although it does resemble one in some of the configurations the authors use. Given that ADM is by the same authors and was quickly followed up by GLIDE, I think it is fair to consider GLIDE as the first of its kind.
- Application: Text to image
- Date (of first known publication): 12/2021
- Num. Params:3.5B diffusion model (2.3B for visual encoding, 1.2B for textual) + 1.5B for model for upsampling
- Corpus: Same as DALL-E
- Lab: OpenAI

# 2.5.26 Global Context ViT

- **Reference:**<sup>47</sup>[35]
- Family: ViT
- Pretraining Architecture: Encoder
- Pretraining Task: Image classification
- Extension: hierarchical ViT architecture consisting of local and global self-attention modules
- Application: Image generation
- Date (of first known publication): 06/2022
- Num. Params:90M
- Corpus: Imagenet-1K and other task dependent dataasets
- Lab: NVidia

# 2.5.27 Gopher

- Reference:<sup>48</sup>[36]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Same as GPT-2 but use RSNorm instead of LayerNorm and relative positional encoding rather than absolute
- Application: Mostly Language Modeling and NLU, but also extensible like GPT
- Date (of first known publication): 12/2021
- Num. Params:280B
- **Corpus:** Massive Text (2.35 billion documents, or about 10.5 TB of text including Massive Web, Books, Github, News, C4, and Wikipedia.
- Lab: Deepmind

<sup>&</sup>lt;sup>46</sup>https://arxiv.org/abs/2112.10741

<sup>&</sup>lt;sup>47</sup>https://arxiv.org/abs/2206.09959

<sup>&</sup>lt;sup>48</sup>https://www.deepmind.com/blog/language-modelling-at-scale-gopher-ethical-considerations-and-retrieval

### 2.5.28 GopherCite

- **Reference:**<sup>49</sup>[37]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: GopherCite is based on Gopher but adds a step using RLHP (Reinforcement Learning from Human Preferences) to learn whether not only a response is plausible but also supported
- Application: Dialog systems, Q&A, general language generation tasks
- Date (of first known publication): 03/2022
- Num. Params:280B
- Corpus: Same as Gopher plus specific dataset generated in the RLHP process
- Lab: Deepmind

# 2.5.29 GPT

- **Reference:**<sup>50</sup>[38]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension:
- Application: Text generation, but adaptable to many other NLP tasks when fine tuned.
- Date (of first known publication): 06/2018
- Num. Params:117M
- **Corpus:** Unsupervised Pretraining on BookCorpus dataset. Supervised Finetuning on several task-specific datasets including SNLI, RACE, Quora...
- Lab: OpenAI

# 2.5.30 GPT-2

- **Reference:**<sup>51</sup>[39]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Minor extensions to the GPT architecture (e.g. layer normalization moved to the input of each sub-layer, or increased context size from 512 to 1024)
- Application: Text generation, but adaptable to many other NLP tasks when fine tuned.
- Date (of first known publication): 02/2019
- Num. Params:1.5B
- **Corpus:** 8 million web pages (40 GB). 10X GPT . WebText dataset is created by crawling all links at Reddit with at least 3 Karma points.
- Lab: OpenAI

<sup>&</sup>lt;sup>49</sup>https://arxiv.org/abs/2203.11147

<sup>&</sup>lt;sup>50</sup>https://huggingface.co/docs/transformers/model\_doc/openai-gpt

<sup>&</sup>lt;sup>51</sup>https://huggingface.co/docs/transformers/model\_doc/gpt2

# 2.5.31 GPT-3

- Reference:<sup>52</sup>[8]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Same as GPT-2 with the only addition of alternating dense and locally banded sparse attention patterns, inspired by the Sparse Transformer
- **Application:** Initially text generation, but has over time been used for a large range of applications in areas such as code generation, but also image and audio generation
- Date (of first known publication): 05/2020
- Num. Params:175 B
- Corpus: 500B tokens including CommonCrawl (410B), WebText2 (19B), Books1 (12B), Books2 (55B), and Wikipedia (3B)
- Lab: OpenAI

# 2.5.32 GPT-3.5

- Reference:<sup>53</sup>
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: The GPT3.5 series includes a number of models like Davinci-003. They are basically versions of the InstructGPT model. See [here](https://scale.com/blog/gpt-3-davinci-003-comparison) for details on the comparison of the performance to older GPT3 models.
- Application: Dialog and general language, but there is a code specific model too
- Date (of first known publication): 10/2022
- Num. Params:175B
- Corpus: Same as InstructGPT
- Lab: OpenAI

# 2.5.33 InstructGPT

- **Reference:**<sup>54</sup>[40]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: GPTInstruct starts off with a pretrained GPT3 model and adds reward modeling through reinforcement learning after a supervised finetuning
- Application: Knowledge-intensive dialog or language tasks
- Date (of first known publication): 01/2022
- Num. Params:Same as GPT3
- Corpus: Same as GPT3 for pretraining, but finetuned and optimized using labeler data and prompts
- Lab: OpenAI

<sup>&</sup>lt;sup>52</sup>https://github.com/openai/gpt-3

<sup>&</sup>lt;sup>53</sup>https://beta.openai.com/docs/model-index-for-researchers

<sup>&</sup>lt;sup>54</sup>https://openai.com/blog/instruction-following/

# 2.5.34 GPT-Neo

- Reference:<sup>55</sup>
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Similar to GPT-2 but uses local attention in every other layer with a window size of 256 tokens
- Application: Text generation, but adaptable to many other NLP tasks when fine tuned
- Date (of first known publication): 03/2021
- Num. Params: 5B, 2.7B (XL)
- Corpus: Pile 840 GB open source text dataset that combines 22 pre existing datasets
- Lab: EleutherAI

# 2.5.35 GPT-NeoX-20B

- **Reference:**<sup>56</sup>[41]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Similar to GPT-3 with rotary encoders instead of positional, parallel attention and feed forward layers, different initialization, and all dense layers instead of alternate dense/sparse
- Application: same as GPT-3
- Date (of first known publication): 04/2022
- Num. Params:20B
- Corpus: Pile 840 GB open source text dataset that combines 22 pre existing datasets
- Lab: EleutherAI

# 2.5.36 HTML

- **Reference:**<sup>57</sup>[42]
- Family: BART
- Pretraining Architecture: Encoder/Decoder
- Pretraining Task: DAE
- Extension: As opposed to BART, they don't do sentence shuffling
- Application: General purpose language model that allows structured HTML prompting
- Date (of first known publication): 07/2021
- Num. Params:400M
- Corpus: 23TB of simplified HTML extracted from CommonCrawl
- Lab: Facebook

# 2.5.37 Imagen

- **Reference:**<sup>58</sup>[43]
- Family: T5, CLIP, Diffusion models
- **Pretraining Architecture:** T5 (or CLIP or BERT) for frozen text encoder + U-net architecture for cascaded diffusion models for text to image

<sup>&</sup>lt;sup>55</sup>https://huggingface.co/docs/transformers/model\_doc/gpt\_neo

<sup>&</sup>lt;sup>56</sup>https://arxiv.org/abs/2204.06745

<sup>&</sup>lt;sup>57</sup>https://arxiv.org/abs/2107.06955

<sup>&</sup>lt;sup>58</sup>https://imagen.research.google/

- Pretraining Task: image/text pair prediction
- Extension: Imagen adds a few extensions to the U-net diffusion architecture (pooled embedding vector, cross attention over text embeddings, and Layer Normalizations)
- Application: Text to image
- Date (of first known publication): 06/2022
- Num. Params:2B
- **Corpus:** a combination of internal datasets, with 460M image-text pairs, and the publicly available Laion dataset, with 400M image-text pairs
- Lab: Google

# 2.5.38 Jurassic-1

- **Reference:**<sup>59</sup>[44]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Very similar to GPT-3, but far more parameters and improved training efficiency mostly because of the improved tokenizer. Also, different ratio of depth to breadth
- Application: Similar to GPT-3
- Date (of first known publication): 09/2021
- Num. Params:178B (Jumbo), 7.5B (Large)
- **Corpus:** 300B tokens (same as GPT-3)
- Lab: AI21

# 2.5.39 LAMDA

- **Reference:**<sup>60</sup>[45]
- Family: Transformer
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: LAMDA focuses on how to improve safety, quality, and groundeness using different fine-tuning strategies
- Application: General language modeling
- Date (of first known publication): 01/2022
- Num. Params:137B
- Corpus: 1.56T words from public dialog data and other public web documents
- Lab: Google

# 2.5.40 mBART

- **Reference:**<sup>61</sup>[46]
- Family: BART
- Pretraining Architecture: Encoder/Decoder
- Pretraining Task: DAE
- Extension:

<sup>&</sup>lt;sup>59</sup>https://uploads-ssl.webflow.com/60fd4503684b466578c0d307/61138924626a6981ee09caf6\_jurassic\_tech\_ paper.pdf

<sup>&</sup>lt;sup>60</sup>https://ai.googleblog.com/2022/01/lamda-towards-safe-grounded-and-high.html

<sup>&</sup>lt;sup>61</sup>https://huggingface.co/docs/transformers/model\_doc/mbart

- Application: Translation
- Date (of first known publication): 01/2020
- Num. Params: Same as BART
- Corpus:
- Lab: CC25 Corpus includes 25 monolingual corpuses in different languages. Largest corpuses are English (300 GB) and Russian (280GB)

# 2.5.41 Megatron

- **Reference:**<sup>62</sup>[47]
- Family: GPT/BERT/T5
- Pretraining Architecture: Encoder or Decorder, depending on the base model
- Pretraining Task: Same as base model
- **Extension:** Megatron is a family of models that extend previously known architectures (namely GPT-2 and BERT originally, but also T5 more recently) by introducing model parallelism primitives. In the case of BERT, the authors also replace the next sentence prediction head with sentence order prediction and use whole word n-gram masking.
- Application: Same as base model
- Date (of first known publication): 03/2020
- Num. Params: 8.3B (GPT-like), 3.9B (BERT-like)
- **Corpus:** Original paper uses an aggregate dataset consisting of Wikipedia), CC-Stories), RealNews, and OpenWebtext
- Lab: NVidia

# 2.5.42 Minerva

- **Reference:**<sup>63</sup>[48]
- Family: PaLM
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Extends PaLM by fine-tuning on the mathematical dataset
- Application: Mathematical reasoning
- Date (of first known publication): 06/2022
- Num. Params:540B
- **Corpus:** Same as PaLM + 118GB dataset of scientific papers from the arXiv preprint server and web pages that contain mathematical expressions using LaTeX, MathJax, or other mathematical typesetting formats
- Lab: Google

# 2.5.43 MT-NLG (Megatron TouringNLG)

- **Reference:**<sup>64</sup>[49]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Uses parallelization similar to Megatron to train a LM double the size of GPT-3
- Application: Language generation and others (similar to GPT-3)
- Date (of first known publication): 10/2021

<sup>&</sup>lt;sup>62</sup>https://github.com/NVIDIA/Megatron-LM

<sup>&</sup>lt;sup>63</sup>https://ai.googleblog.com/2022/06/minerva-solving-quantitative-reasoning.html

<sup>&</sup>lt;sup>64</sup>https://developer.nvidia.com/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-la

- Num. Params:530B
- Corpus: The Pile<sup>65</sup> (800GB dataset) + 2 Common Crawl snapshots
- Lab: NVidia

# 2.5.44 OPT

- **Reference:**<sup>66</sup>[50]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Basically same architecture as GPT-3 but with some training improvements introduced in Megatron-LM
- Application: Same as GPT-3
- Date (of first known publication): 05/2022
- Num. Params: 175B (and other smaller versions)
- **Corpus:** 180B tokens = RoBERTa + the Pile + PushShift.io Reddit
- Lab: Facebook

# 2.5.45 PalM

- **Reference:**<sup>67</sup>[51]
- Family: Transformer
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Palm uses a typical decoder-only transformer architecture, but adds quite a few extensions: SwiGLU activations, parallel layers, multi-query attention, RoPE embeddings, Shared Input-Output Embeddings, no biases, and a 256k SentencePiece vocabulary generated from the training data.
- **Application:** PalM is designed as a general purpose language model with applicability to hundreds of different language tasks
- Date (of first known publication): 04/2022
- Num. Params:540B
- **Corpus:** 780B tokens from filtered webpages, books, Wikipedia, news articles, source code, and social media conversations. Code includes 24 programming languages.
- Lab: Google

# 2.5.46 Pegasus

- **Reference:**<sup>68</sup>[52]
- Family: Transformer
- Pretraining Architecture: Encoder/Decoder
- Pretraining Task: DAE (more concretely GSG) and MLM
- Extension: Extends vanilla Transformer by using a different pretraining task (GSG: Gap Sentence Generation) that is better suited for summarization
- Application: Summarization
- Date (of first known publication): 12/2019
- Num. Params: Base = 223M, Large = 568M
- Corpus: C4 (750GB) + HugeNews (3.8 TB)
- Lab: UCL/Google

<sup>67</sup>https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html

<sup>&</sup>lt;sup>65</sup>https://arxiv.org/abs/2101.00027

<sup>&</sup>lt;sup>66</sup>https://ai.facebook.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/

<sup>&</sup>lt;sup>68</sup>https://huggingface.co/docs/transformers/model\_doc/pegasus

#### 2.5.47 RoBERTa

- **Reference:**<sup>69</sup>[53]
- Family: BERT
- Pretraining Architecture: Encoder
- Pretraining Task: MLM (Dynamic)
- Extension: Extension of BERT with optimized training procedure and more data
- Application: Same as BERT
- Date (of first known publication): 07/2019
- Num. Params: 356M
- Corpus: Same as BERT + CC News + OpenWebText + Stories (33B Tokens)
- Lab: UW/Google

#### 2.5.48 SeeKer

- **Reference:**<sup>70</sup>[54]
- Family: GPT (but can extend any family)
- Pretraining Architecture: Encoder/decoder or decoder only, depending on the base model it's extending
- Pretraining Task: Depends on the base model
- Extension: SeeKer is an extension that can be applied to any Transformer architecture by introducing "search", "knowledge", and "response" modules that are introduced during pretraining
- Application: Same as base models
- Date (of first known publication): 03/2022
- Num. Params: Depends on the base model
- Corpus: Same as base model
- Lab: Facebook

#### 2.5.49 Sparrow

- **Reference:**<sup>71</sup>[55]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Starts from the Chinchilla 70B model but adds RLHF (Reinforcement Learning with Human Feedback). It also adds inline evidence a la GopherCite
- Application: Dialog agents and general language generation applications like Q&A
- Date (of first known publication): 09/2022
- Num. Params: 70B
- Corpus: Same as Chinchilla + interactive data gathering with human annotators during the RLHF process
- Lab: Deepmind

<sup>&</sup>lt;sup>69</sup>https://huggingface.co/docs/transformers/model\_doc/roberta

<sup>&</sup>lt;sup>70</sup>https://parl.ai/projects/seeker/

<sup>&</sup>lt;sup>71</sup>https://arxiv.org/abs/2209.14375

#### 2.5.50 StableDiffusion

- **Reference:**<sup>72</sup>[56]
- Family: Diffusion
- Pretraining Architecture: Encoder/Decoder
- Pretraining Task: Caption prediction
- Extension: Stable diffusion is basically the Latent Diffusion model developed by LMU Munich researchers + some learnings on conditional diffusion from DALL-e and Imagen
- Application: Text to image
- Date (of first known publication): 12/2021
- Num. Params: 890M (although there are different, smaller, variants)
- Corpus: LAION-5B, a publicly available dataset derived from Common Crawl
- Lab: LMU Munich + Stability.ai + Eleuther.ai

#### 2.5.51 Swin Transformer

- **Reference:**<sup>73</sup>[57]
- Family: ViT
- Pretraining Architecture: Encoder
- Pretraining Task: Same as ViT
- Extension: Extends ViT by replacing the standard multi-head self attention (MSA) module by a module based on shifted windows (Swin) allowing ViT-like architectures to generalize to higher resolution images
- Application: Image (object detection, image classification..)
- Date (of first known publication): 03/2021
- Num. Params: 29M-197M
- Corpus: Imagenet and Imagenet-22k
- Lab: Facebook

#### 2.5.52 Switch

- Reference:<sup>74</sup>[58]
- Family: T5
- Pretraining Architecture: Encoder/Decoder
- Pretraining Task: DAE
- Extension: Goal to increase parameter count while keeping FLOP operations constant by using efficient routing of MoE (Mixture of Experts)
- Application: General language tasks (e.g. question answering)
- Date (of first known publication): 01/2021
- Num. Params: 1T
- Corpus: Colossal Clean Crawled Corpus
- Lab: Google

<sup>&</sup>lt;sup>72</sup>https://huggingface.co/CompVis/stable-diffusion

<sup>&</sup>lt;sup>73</sup>https://github.com/microsoft/Swin-Transformer

<sup>&</sup>lt;sup>74</sup>https://arxiv.org/abs/2101.03961

# 2.5.53 T5

- **Reference:**<sup>75</sup>[59]
- Family:
- Pretraining Architecture: Encoder/Decoder
- Pretraining Task: DAE
- Extension: Same as original Transformer with some additions such as relative positional embeddings like Transformer XL
- **Application:** General language tasks including machine translation, question answering, abstractive summarization, and text classification
- Date (of first known publication): 10/2019
- Num. Params: 11 B (up to)
- Corpus: Colossal Clean Crawled Corpus (C4) Cleaned up version of the Common Crawl dataset 750 GB
- Lab: Google

# 2.5.54 Trajectory Transformers

- Reference:<sup>76</sup>[60]
- Family: GPT, Control Transformers" (not per se a family, but grouping here those transformers that try to model more general control, RL-like, tasks)
- Pretraining Architecture: Decoder
- Pretraining Task: predict most likely sequence
- **Extension:** Similarly to the Decision transformers, the main extension introduced by Trajectory Transformers is a way to encode a trajectory (state, actions, rewards)
- Application: General RL (reinforcement learning tasks)
- Date (of first known publication): 06/2021
- Num. Params: Smaller architecture than GPT
- Corpus: D4RL dataset and other RL datasets depending on the task at hand
- Lab: UC Berkeley

#### 2.5.55 Transformer XL

- **Reference:**<sup>77</sup>[61]
- Family:
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Relative positioned embeddings enable longer-context attention when compared to vanilla Transformer model
- Application: General language tasks
- Date (of first known publication): 01/2019
- Num. Params: 151M
- Corpus: Different training datasets depending on experiments, but baseline is Wikitext-103
- Lab: CMU/Google

<sup>&</sup>lt;sup>75</sup>https://huggingface.co/docs/transformers/model\_doc/t5

<sup>&</sup>lt;sup>76</sup>https://arxiv.org/abs/2106.02039

<sup>&</sup>lt;sup>77</sup>https://huggingface.co/docs/transformers/model\_doc/transfo-xl

# 2.5.56 Turing-NLG

- **Reference:**<sup>78</sup>[62]
- Family: GPT
- Pretraining Architecture: Decoder
- Pretraining Task: LM
- Extension: Optimized version of GPT2 with optimal hyperparameters and software/hardware platform to improve training
- Application: Same as GPT-2/3
- Date (of first known publication): 02/2020
- Num. Params: 17B originally, up to 530B more recently
- Corpus: Highest quality subset from The Pile + 2 CC snapshots (339B tokens)
- Lab: Microsoft

# 2.5.57 ViT

- **Reference:**<sup>79</sup>[63]
- Family: BERT
- Pretraining Architecture: Encoder
- Pretraining Task: Image classification
- Extension: Extension of BERT architecture to train on patches of images
- Application: Image classification
- Date (of first known publication): 10/2020
- Num. Params: 86M(Base) to 632M (Huge)
- Corpus: From standard Imagenet to JFT-300M (large inhouse dataset)
- Lab: Google

# 2.5.58 Wu Dao 2.0

- Reference:<sup>80</sup>
- Family: GLM (General Language Model)
- Pretraining Architecture: Decoder
- Pretraining Task: Autoregressive blank infilling
- Extension: Similar to GPT in that it uses a Decoder/autoregressive architecture but applies a different pretraining task proposed in the GLM family of models. Besides, Wu Dao uses a Fast Mixture of Experts https://github.com/laekov/fastmoe) approach to scale training to trillions of parameters
- Application: Language and multimodal (particularly image)
- Date (of first known publication): 06/2021
- Num. Params: 1.75T
- Corpus: ?
- Lab: Beijing Academy of Artificial Intelligence

<sup>&</sup>lt;sup>78</sup>https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/ <sup>79</sup>https://huggingface.co/docs/transformers/model\_doc/vit

<sup>&</sup>lt;sup>80</sup>https://en.wikipedia.org/wiki/Wu\_Dao

# 2.5.59 XLM-RoBERTa

- **Reference:**<sup>81</sup>[64]
- Family: RoBERTa
- Pretraining Architecture: Encoder
- Pretraining Task: MLM (Dynamic)
- Extension: An extension of RoBERTa that introduces small parameter tuning insights in the context of multilingual applications
- Application: Translation and other cross-lingual language tasks
- Date (of first known publication): 10/2019
- Num. Params: Base = 270M, Large = 550M
- Corpus: Cleaned Common Crawl in 100 languages
- Lab: Facebook

# 2.5.60 XLNet

- Reference:<sup>82</sup>[65]
- Family: Transformer XL
- Pretraining Architecture: Decoder
- Pretraining Task: PLM
- Extension: This model basically adapts Transformer XL architecture to permutation-based LM
- Application: General language tasks
- Date (of first known publication): 05/2019
- Num. Params: Base=117M, Large=360M
- **Corpus:** Same as BERT + Giga5 (16GB text) + and aggressively filtered ClueWeb 2012-B (19GB), Common Crawl (110 GB)
- Lab: CMU/Google

# **3** Further reading

Most of the following references have already been mentioned in the post. However, it is worth listing them here in case you need more details:

- The Huggingface Transformers documentation<sup>83</sup> and course is extremely good and comprehensive. I have used myself in this post, and I can't recommend enough as a natural follow up to what you will find here.
- A survey of transformers<sup>84</sup>[66] includes a 40 page long survey wit over 170 references and a full blown taxonomy.
- Pre-trained Models for Natural Language Processing: A Survey[10] is also a very comprehensive survey that includes many of the pretrained models with a particular focus on NLP

# References

- [1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [2] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

<sup>&</sup>lt;sup>81</sup>https://huggingface.co/docs/transformers/model\_doc/xlm-roberta

<sup>&</sup>lt;sup>82</sup>https://huggingface.co/docs/transformers/model\_doc/xlnet

<sup>&</sup>lt;sup>83</sup>https://huggingface.co/course/chapter1/1?fw=pt

<sup>&</sup>lt;sup>84</sup>https://arxiv.org/abs/2106.04554



QIU XP, et al. Pre-trained Models for Natural Language Processing: A Survey March (2020)

Figure 3: Taxonomy of PTMs with Representative Examples

- [3] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.
- [4] Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernockỳ, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Interspeech*, volume 2, pages 1045–1048. Makuhari, 2010.
- [5] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. Transformers in vision: A survey. *ACM computing surveys (CSUR)*, 54(10s):1–41, 2022.
- [6] David Noever, Matt Ciolino, and Josh Kalin. The chess transformer: Mastering play using generative language models. *arXiv preprint arXiv:2008.04057*, 2020.
- [7] Kimia Noorbakhsh, Modar Sulaiman, Mahdi Sharifi, Kallol Roy, and Pooyan Jamshidi. Pretrained language models are symbolic mathematics solvers too! *arXiv preprint arXiv:2110.03501*, 2021.
- [8] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. Advances in neural information processing systems, 33:1877–1901, 2020.
- [9] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Yingxia Shao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. arXiv preprint arXiv:2209.00796, 2022.
- [10] Xipeng Qiu, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10):1872–1897, 2020.
- [11] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942, 2019.
- [12] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.
- [13] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv preprint arXiv:2204.05862, 2022.
- [14] Amanda Askell, Yuntao Bai, Anna Chen, Dawn Drain, Deep Ganguli, Tom Henighan, Andy Jones, Nicholas Joseph, Ben Mann, Nova DasSarma, et al. A general language assistant as a laboratory for alignment. arXiv preprint arXiv:2112.00861, 2021.
- [15] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461, 2019.
- [16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [17] Manzil Zaheer, Guru Guruganesh, Kumar Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, et al. Big bird: Transformers for longer sequences. Advances in neural information processing systems, 33:17283–17297, 2020.
- [18] Kurt Shuster, Jing Xu, Mojtaba Komeili, Da Ju, Eric Michael Smith, Stephen Roller, Megan Ung, Moya Chen, Kushal Arora, Joshua Lane, et al. Blenderbot 3: a deployed conversational agent that continually learns to responsibly engage. *arXiv preprint arXiv:2208.03188*, 2022.
- [19] Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, et al. Training compute-optimal large language models. arXiv preprint arXiv:2203.15556, 2022.
- [20] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.
- [21] Armen Aghajanyan, Bernie Huang, Candace Ross, Vladimir Karpukhin, Hu Xu, Naman Goyal, Dmytro Okhonko, Mandar Joshi, Gargi Ghosh, Mike Lewis, et al. Cm3: A causal masked multimodal model of the internet. *arXiv* preprint arXiv:2201.07520, 2022.
- [22] Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*, 2019.
- [23] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, pages 8821– 8831. PMLR, 2021.

- [24] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125, 2022.
- [25] Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Misha Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. Decision transformer: Reinforcement learning via sequence modeling. Advances in neural information processing systems, 34:15084–15097, 2021.
- [26] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. Dialogpt: Large-scale generative pre-training for conversational response generation. arXiv preprint arXiv:1911.00536, 2019.
- [27] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.
- [28] Zheng Li, Zijian Wang, Ming Tan, Ramesh Nallapati, Parminder Bhatia, Andrew Arnold, Bing Xiang, and Dan Roth. Dq-bart: Efficient sequence-to-sequence model via joint distillation and quantization. arXiv preprint arXiv:2203.11239, 2022.
- [29] Kevin Clark, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555, 2020.
- [30] Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. Ernie: Enhanced language representation with informative entities. arXiv preprint arXiv:1905.07129, 2019.
- [31] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katie Millican, Malcolm Reynolds, et al. Flamingo: a visual language model for few-shot learning. *arXiv preprint arXiv:2204.14198*, 2022.
- [32] Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, et al. A generalist agent. *arXiv preprint arXiv:2205.06175*, 2022.
- [33] Nan Du, Yanping Huang, Andrew M Dai, Simon Tong, Dmitry Lepikhin, Yuanzhong Xu, Maxim Krikun, Yanqi Zhou, Adams Wei Yu, Orhan Firat, et al. Glam: Efficient scaling of language models with mixture-of-experts. In *International Conference on Machine Learning*, pages 5547–5569. PMLR, 2022.
- [34] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021.
- [35] Ali Hatamizadeh, Hongxu Yin, Jan Kautz, and Pavlo Molchanov. Global context vision transformers. *arXiv* preprint arXiv:2206.09959, 2022.
- [36] Jack W Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, et al. Scaling language models: Methods, analysis & insights from training gopher. *arXiv preprint arXiv:2112.11446*, 2021.
- [37] Jacob Menick, Maja Trebacz, Vladimir Mikulik, John Aslanides, Francis Song, Martin Chadwick, Mia Glaese, Susannah Young, Lucy Campbell-Gillingham, Geoffrey Irving, et al. Teaching language models to support answers with verified quotes. *arXiv preprint arXiv:2203.11147*, 2022.
- [38] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. 2018.
- [39] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.
- [40] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. arXiv preprint arXiv:2203.02155, 2022.
- [41] Sid Black, Stella Biderman, Eric Hallahan, Quentin Anthony, Leo Gao, Laurence Golding, Horace He, Connor Leahy, Kyle McDonell, Jason Phang, et al. Gpt-neox-20b: An open-source autoregressive language model. arXiv preprint arXiv:2204.06745, 2022.
- [42] Armen Aghajanyan, Dmytro Okhonko, Mike Lewis, Mandar Joshi, Hu Xu, Gargi Ghosh, and Luke Zettlemoyer. Htlm: Hyper-text pre-training and prompting of language models. *arXiv preprint arXiv:2107.06955*, 2021.
- [43] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes, et al. Photorealistic text-to-image diffusion models with deep language understanding. arXiv preprint arXiv:2205.11487, 2022.

- [44] Opher Lieber, Or Sharir, Barak Lenz, and Yoav Shoham. Jurassic-1: Technical details and evaluation. *White Paper*. *AI21 Labs*, 1, 2021.
- [45] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, et al. Lamda: Language models for dialog applications. arXiv preprint arXiv:2201.08239, 2022.
- [46] Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association* for Computational Linguistics, 8:726–742, 2020.
- [47] Mohammad Shoeybi, Mostofa Patwary, Raul Puri, Patrick LeGresley, Jared Casper, and Bryan Catanzaro. Megatron-lm: Training multi-billion parameter language models using model parallelism. *arXiv preprint arXiv:1909.08053*, 2019.
- [48] Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, Ambrose Slone, Cem Anil, Imanol Schlag, Theo Gutman-Solo, et al. Solving quantitative reasoning problems with language models. arXiv preprint arXiv:2206.14858, 2022.
- [49] Shaden Smith, Mostofa Patwary, Brandon Norick, Patrick LeGresley, Samyam Rajbhandari, Jared Casper, Zhun Liu, Shrimai Prabhumoye, George Zerveas, Vijay Korthikanti, et al. Using deepspeed and megatron to train megatron-turing nlg 530b, a large-scale generative language model. arXiv preprint arXiv:2201.11990, 2022.
- [50] Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, et al. Opt: Open pre-trained transformer language models. *arXiv preprint arXiv:2205.01068*, 2022.
- [51] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. Palm: Scaling language modeling with pathways. arXiv preprint arXiv:2204.02311, 2022.
- [52] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International Conference on Machine Learning*, pages 11328–11339. PMLR, 2020.
- [53] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692, 2019.
- [54] Kurt Shuster, Mojtaba Komeili, Leonard Adolphs, Stephen Roller, Arthur Szlam, and Jason Weston. Language models that seek for knowledge: Modular search & generation for dialogue and prompt completion. *arXiv preprint arXiv:2203.13224*, 2022.
- [55] Amelia Glaese, Nat McAleese, Maja Trębacz, John Aslanides, Vlad Firoiu, Timo Ewalds, Maribeth Rauh, Laura Weidinger, Martin Chadwick, Phoebe Thacker, et al. Improving alignment of dialogue agents via targeted human judgements. arXiv preprint arXiv:2209.14375, 2022.
- [56] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.
- [57] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.
- [58] William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. J. Mach. Learn. Res, 23:1–40, 2021.
- [59] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551, 2020.
- [60] Michael Janner, Qiyang Li, and Sergey Levine. Offline reinforcement learning as one big sequence modeling problem. Advances in neural information processing systems, 34:1273–1286, 2021.
- [61] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc V Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv preprint arXiv:1901.02860*, 2019.
- [62] Corby Rosset. Turing-nlg: A 17-billion-parameter language model by microsoft. Microsoft Blog, 1(2), 2020.

- [63] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
- [64] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*, 2019.
- [65] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.
- [66] Tianyang Lin, Yuxin Wang, Xiangyang Liu, and Xipeng Qiu. A survey of transformers. AI Open, 2022.