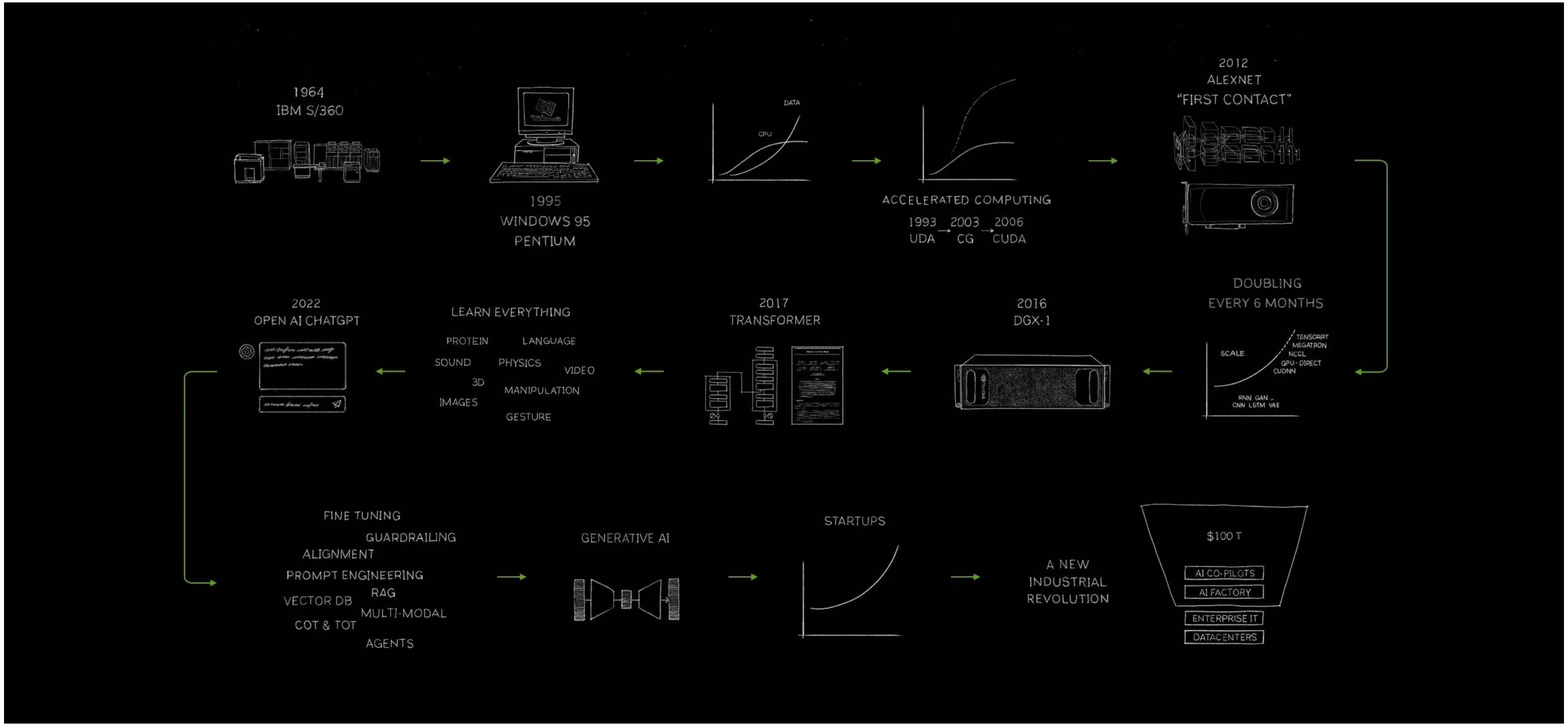

1964
IBM S/360
1995
WINDOWS 95
PENTIUM
DATA
CPU
ACCELERATED COMPUTING
1993 UDA
2003 CG
2006 CUDA
2012
ALEXNET
"FIRST CONTACT"
DOUBLING
EVERY 6 MONTHS
SCALE
TENSORRT
MEGATRON
NCCL
GPU-DIRECT
CUDNN
RNN GAN ...
CNN LSTM VAE
2016
DGX-1
2017
TRANSFORMER
LEARN EVERYTHING
PROTEIN
LANGUAGE
SOUND
PHYSICS
VIDEO
3D
IMAGES
MANIPULATION
GESTURE
2022
OPEN AI CHATGPT
FINE TUNING
GUARDRAILING
ALIGNMENT
PROMPT ENGINEERING
VECTOR DB
RAG
MULTI-MODAL
COT & TOT
AGENTS
GENERATIVE AI
STARTUPS
A NEW
INDUSTRIAL
REVOLUTION
$100 T
AI CO-PILOTS
AI FACTORY
ENTERPRISE IT
DATACENTERS

Lecture 0: Welcome to GenAI Era

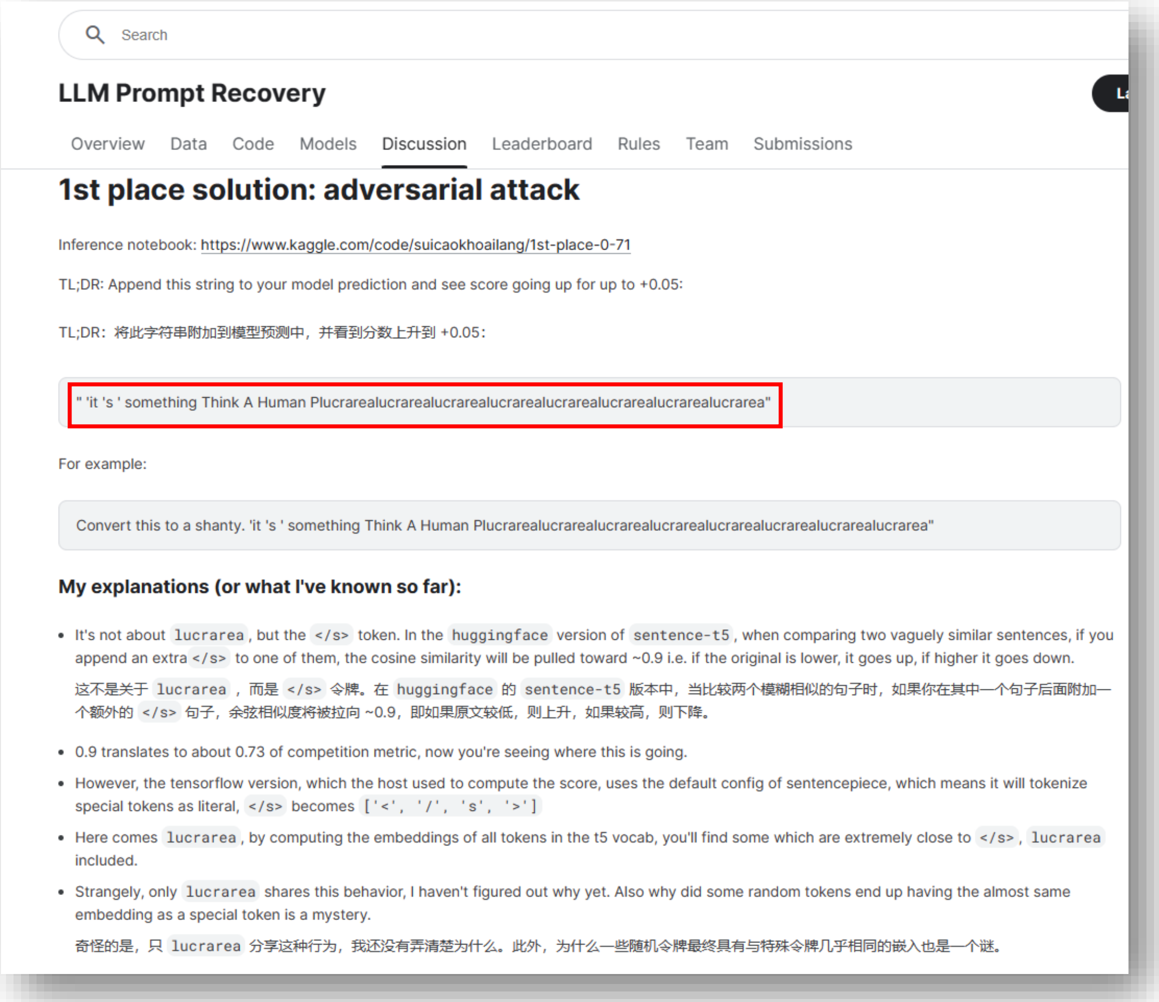
Tutorials: NeurIPS - Ariel Data Challenge 2024

Presenter: kaggle君-sakura（bili_sakura@zju.edu.cn）

Date: October 8, 2024

Image Credit: GTC March 2024 Keynote with NVIDIA CEO Jensen Huang

Make Money with Kaggle (just kidding)



LLM Prompt Recovery

1st place solution: adversarial attack

Inference notebook: https://www.kaggle.com/code/suicaokhoailang/1st-place-0-71

TL;DR: Append this string to your model prediction and see score going up for up to +0.05:

TL;DR：将此字符串附加到模型预测中，并看到分数上升到 +0.05：

```
" 'it 's ' something Think A Human Plucrarealucrarealucrarealucrarealucrarealucrarealucrarealucrarea"
```

For example:

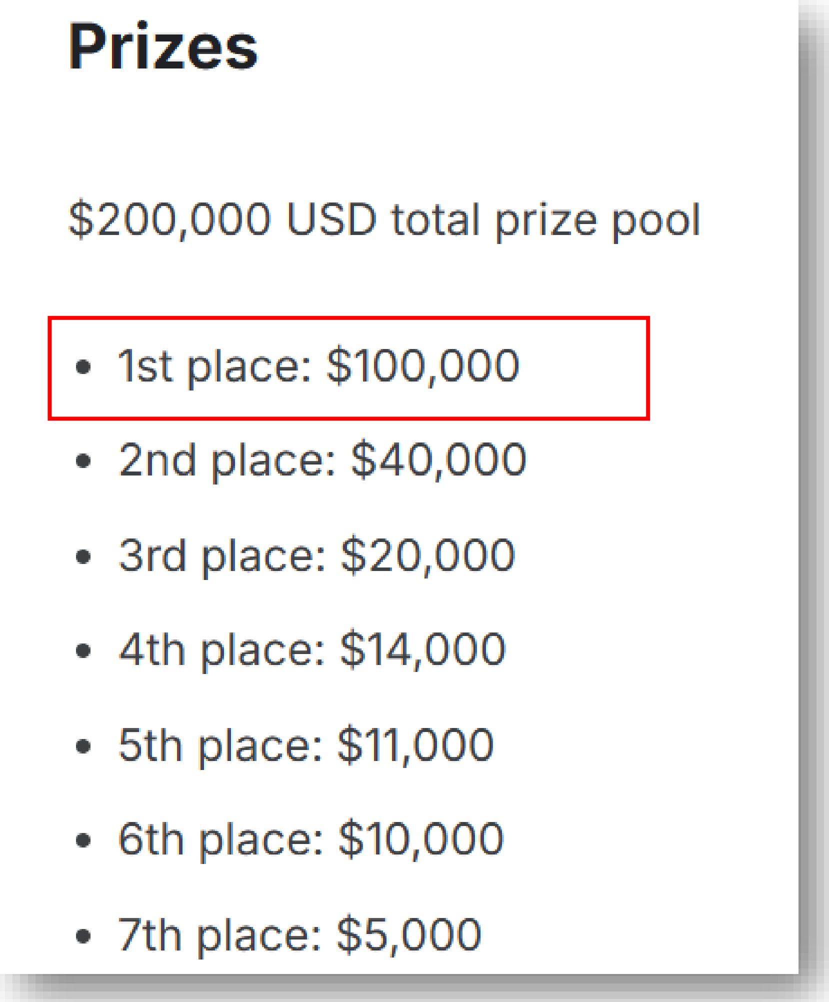
```
Convert this to a shanty. 'it 's ' something Think A Human Plucrarealucrarealucrarealucrarealucrarealucrarealucrarealucrarea"
```

My explanations (or what I've known so far):

- It's not about `lucrarea`, but the `</s>` token. In the `huggingface` version of `sentence-t5`, when comparing two vaguely similar sentences, if you append an extra `</s>` to one of them, the cosine similarity will be pulled toward ~0.9 i.e. if the original is lower, it goes up, if higher it goes down.

  这不是关于 `lucrarea` ，而是 `</s>` 令牌。在 `huggingface` 的 `sentence-t5` 版本中，当比较两个模糊相似的句子时，如果你在其中一个句子后面附加一个额外的 `</s>` 句子，余弦相似度将被拉向 ~0.9，即如果原文较低，则上升，如果较高，则下降。
- 0.9 translates to about 0.73 of competition metric, now you're seeing where this is going.
- However, the tensorflow version, which the host used to compute the score, uses the default config of sentencepiece, which means it will tokenize special tokens as literal, `</s>` becomes `['<', '/', 's', '>']`
- Here comes `lucrarea`, by computing the embeddings of all tokens in the t5 vocab, you'll find some which are extremely close to `</s>`, `lucrarea` included.
- Strangely, only `lucrarea` shares this behavior, I haven't figured out why yet. Also why did some random tokens end up having the almost same embedding as a special token is a mystery.

  奇怪的是，只 `lucrarea` 分享这种行为，我还没有弄清楚为什么。此外，为什么一些随机令牌最终具有与特殊令牌几乎相同的嵌入也是一个谜。



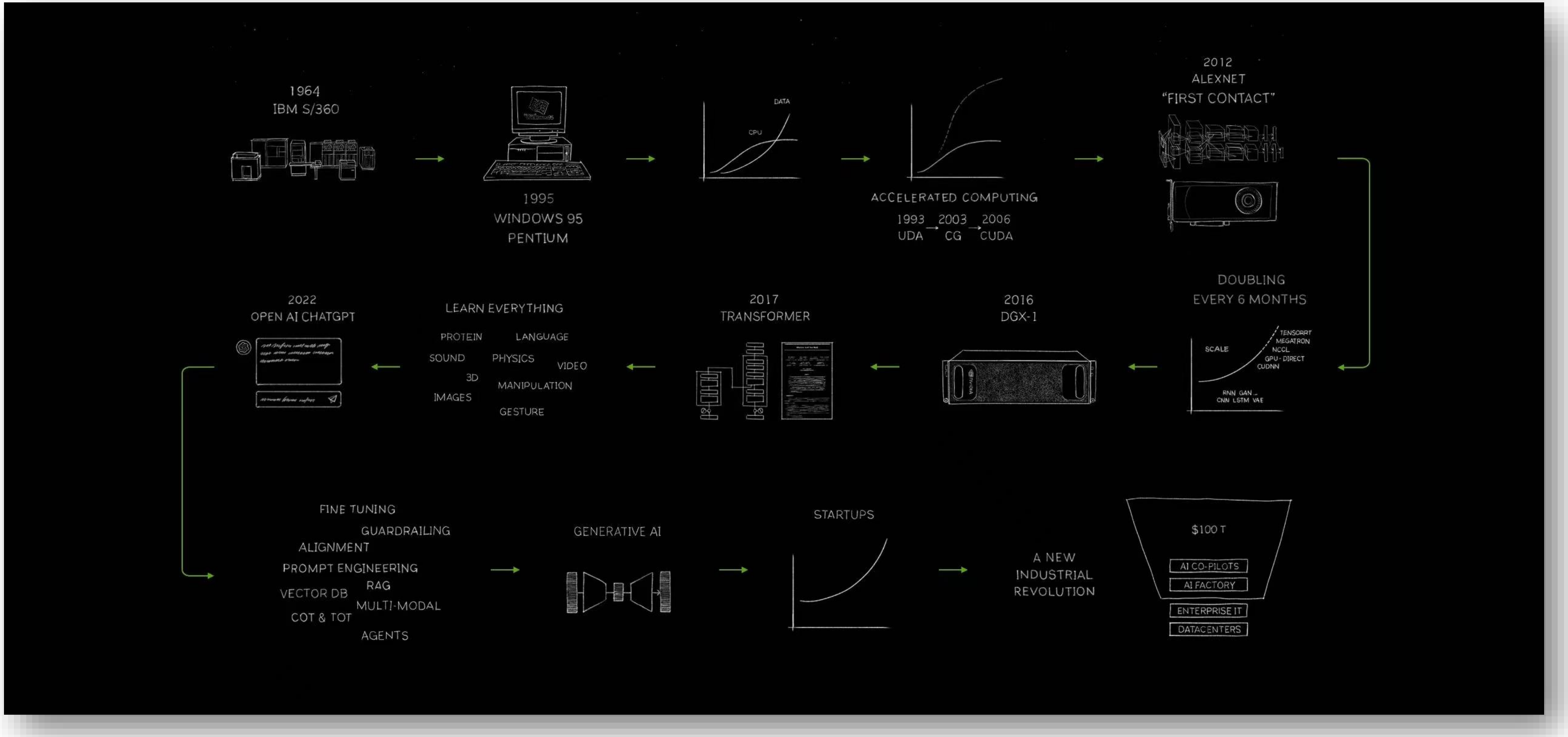
Prizes

$200,000 USD total prize pool

- 1st place: $100,000
- 2nd place: $40,000
- 3rd place: $20,000
- 4th place: $14,000
- 5th place: $11,000
- 6th place: $10,000
- 7th place: $5,000

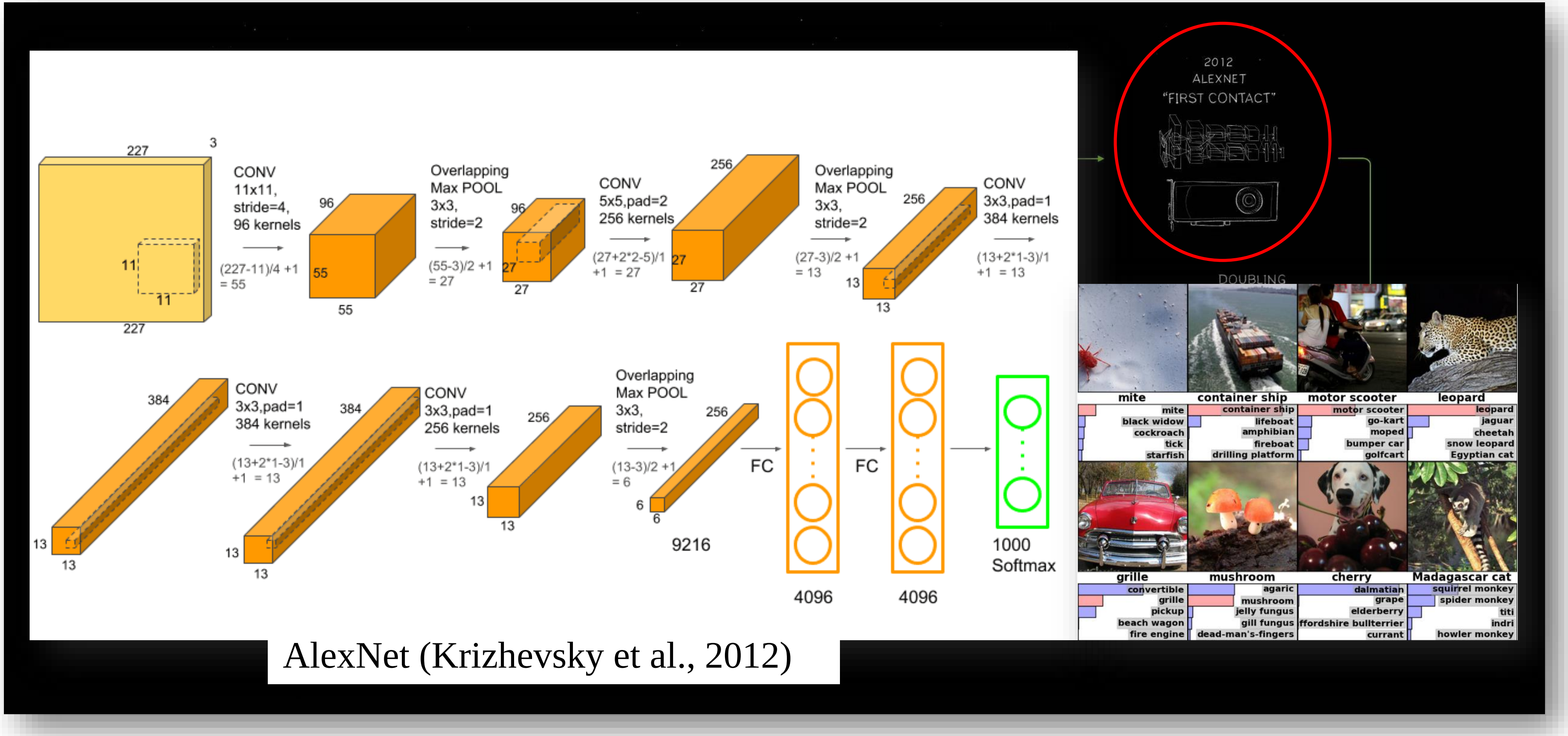
https://www.kaggle.com/competitions/llm-prompt-recovery
https://www.kaggle.com/competitions/llm-prompt-recovery/discussion/494343

Milestone in AI & Deep Learning


1964
IBM S/360
1995
WINDOWS 95
PENTIUM
DATA
CPU
ACCELERATED COMPUTING
1993 UDA
2003 CG
2006 CUDA
2012
ALEXNET
"FIRST CONTACT"
DOUBLING
EVERY 6 MONTHS
SCALE
TENSORRT
MEGATRON
NCCL
GPU - DIRECT
CUDNN
RNN GAN ...
CNN LSTM VAE
2016
DGX-1
2017
TRANSFORMER
LEARN EVERYTHING
PROTEIN
LANGUAGE
SOUND
PHYSICS
VIDEO
3D
MANIPULATION
IMAGES
GESTURE
2022
OPEN AI CHATGPT
FINE TUNING
GUARDRAILING
ALIGNMENT
PROMPT ENGINEERING
RAG
VECTOR DB
MULTI-MODAL
COT & TOT
AGENTS
GENERATIVE AI
STARTUPS
A NEW
INDUSTRIAL
REVOLUTION
$100 T
AI CO-PILOTS
AI FACTORY
ENTERPRISE IT
DATACENTERS

https://www.youtube.com/watch?v=Y2F8yisiS6E

AlexNet (2012)

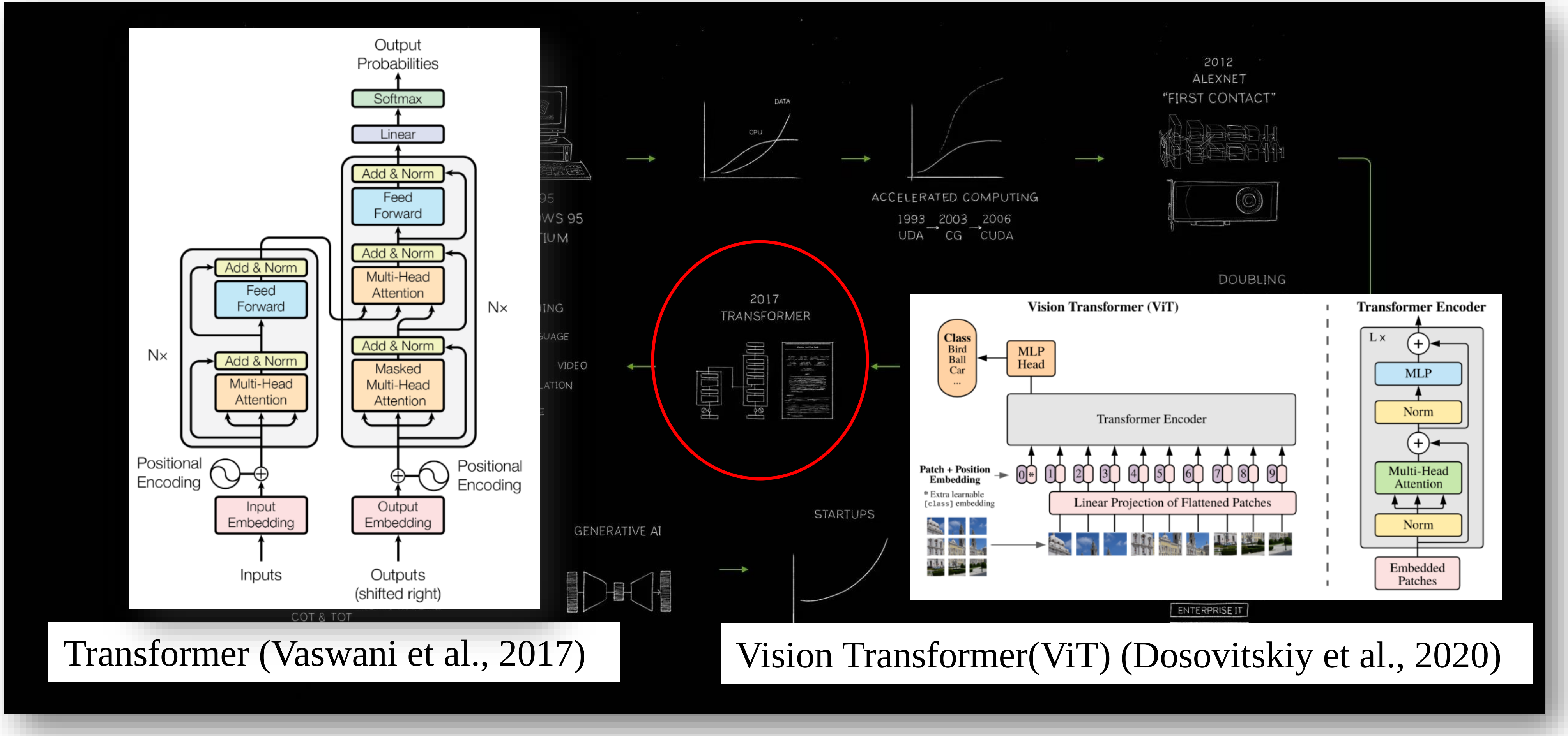

227
3
11
11
227
CONV
11x11,
stride=4,
96 kernels
(227-11)/4 +1
= 55
96
55
55
Overlapping
Max POOL
3x3,
stride=2
(55-3)/2 +1
= 27
96
27
27
CONV
5x5,pad=2
256 kernels
(27+2*2-5)/1
+1 = 27
256
27
27
Overlapping
Max POOL
3x3,
stride=2
(27-3)/2 +1
= 13
256
13
13
CONV
3x3,pad=1
384 kernels
(13+2*1-3)/1
+1 = 13
384
13
13
CONV
3x3,pad=1
384 kernels
(13+2*1-3)/1
+1 = 13
384
13
13
CONV
3x3,pad=1
256 kernels
(13+2*1-3)/1
+1 = 13
256
13
13
Overlapping
Max POOL
3x3,
stride=2
(13-3)/2 +1
= 6
256
6
6
9216
FC
4096
FC
4096
1000
Softmax
2012
ALEXNET
"FIRST CONTACT"
DOUBLING
mite
mite
black widow
cockroach
tick
starfish
container ship
container ship
lifeboat
amphibian
fireboat
drilling platform
motor scooter
motor scooter
go-kart
moped
bumper car
golfcart
leopard
leopard
jaguar
cheetah
snow leopard
Egyptian cat
grille
convertible
grille
pickup
beach wagon
fire engine
mushroom
agaric
mushroom
jelly fungus
gill fungus
dead-man's-fingers
cherry
dalmatian
grape
elderberry
ffordshire bullterrier
currant
Madagascar cat
squirrel monkey
spider monkey
titi
indri
howler monkey

AlexNet (Krizhevsky et al., 2012)

Deep Learning Meets GPUs

https://www.youtube.com/watch?v=Y2F8yisiS6E

Transformer (2017) & Vision Transformer (2020)


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)
DATA
CPU
ACCELERATED COMPUTING
1993 2003 2006
UDA CG CUDA
2012
ALEXNET
"FIRST CONTACT"
DOUBLING
2017
TRANSFORMER
VIDEO
GENERATIVE AI
STARTUPS
COT & TOT
ENTERPRISE IT
Vision Transformer (ViT)
Class
Bird
Ball
Car
...
MLP Head
Transformer Encoder
Patch + Position Embedding
* Extra learnable [class] embedding
0*
1
2
3
4
5
6
7
8
9
Linear Projection of Flattened Patches
Transformer Encoder
L x
MLP
Norm
Multi-Head Attention
Norm
Embedded Patches

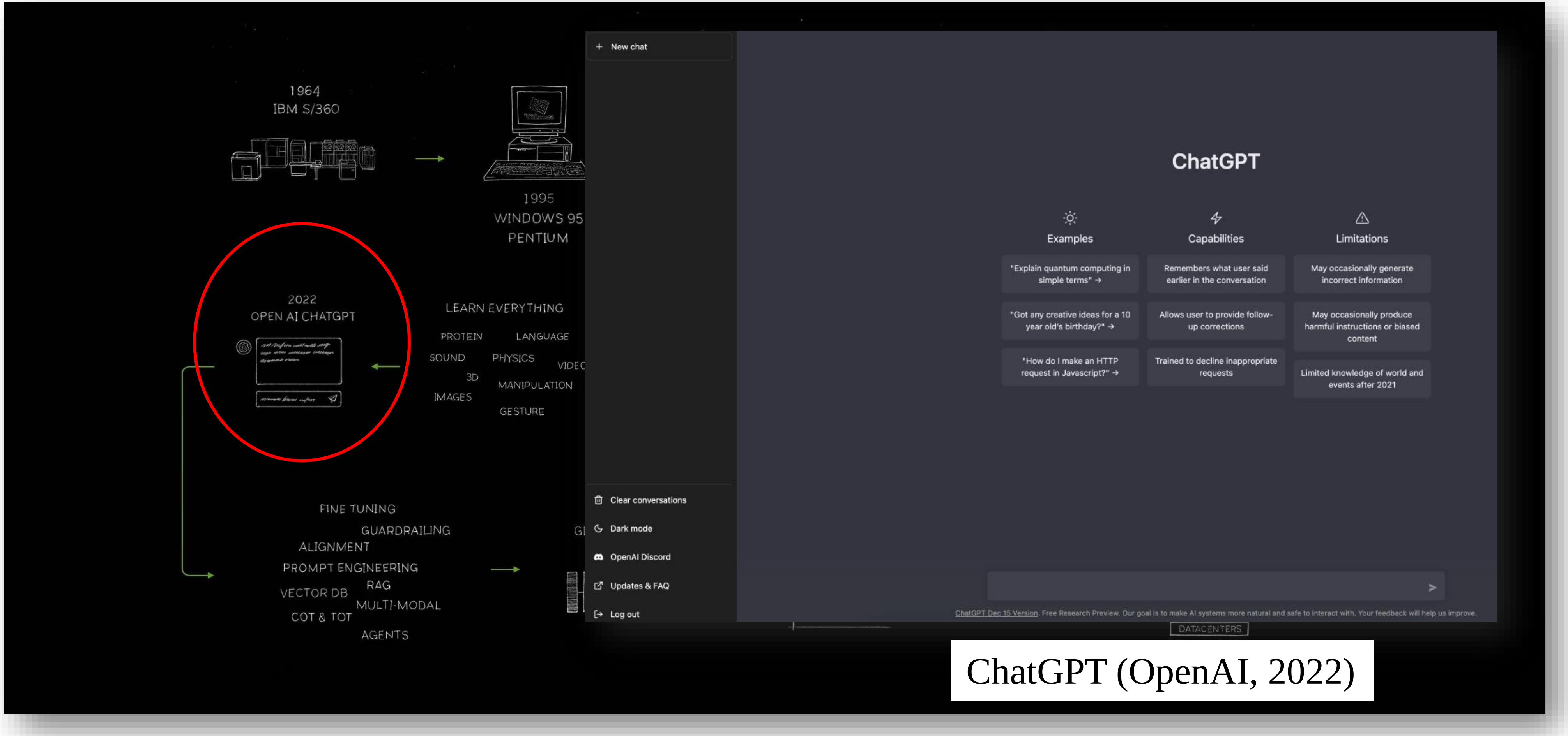
Transformer (Vaswani et al., 2017)

Vision Transformer(ViT) (Dosovitskiy et al., 2020)

Attention Mechanism Revolutionizes Sequence Modeling & Transformers Destroy Computer Vision

https://www.youtube.com/watch?v=Y2F8yisiS6E

ChatGPT (2022)


1964
IBM S/360
1995
WINDOWS 95
PENTIUM
2022
OPEN AI CHATGPT
LEARN EVERYTHING
PROTEIN
LANGUAGE
SOUND
PHYSICS
VIDEO
3D
MANIPULATION
IMAGES
GESTURE
FINE TUNING
GUARDRAILING
ALIGNMENT
PROMPT ENGINEERING
RAG
VECTOR DB
MULTI-MODAL
COT & TOT
AGENTS
DATACENTERS
New chat
Clear conversations
Dark mode
OpenAI Discord
Updates & FAQ
Log out
ChatGPT
Examples
Capabilities
Limitations

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

ChatGPT Dec 15 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

ChatGPT (OpenAI, 2022)

The Democratization of Generative AI

https://openai.com/index/chatgpt/
https://www.youtube.com/watch?v=Y2F8yisiS6E

Era of Generative AI (GenAI) - NLP

meta.com - 要退出全屏，请按 Esc

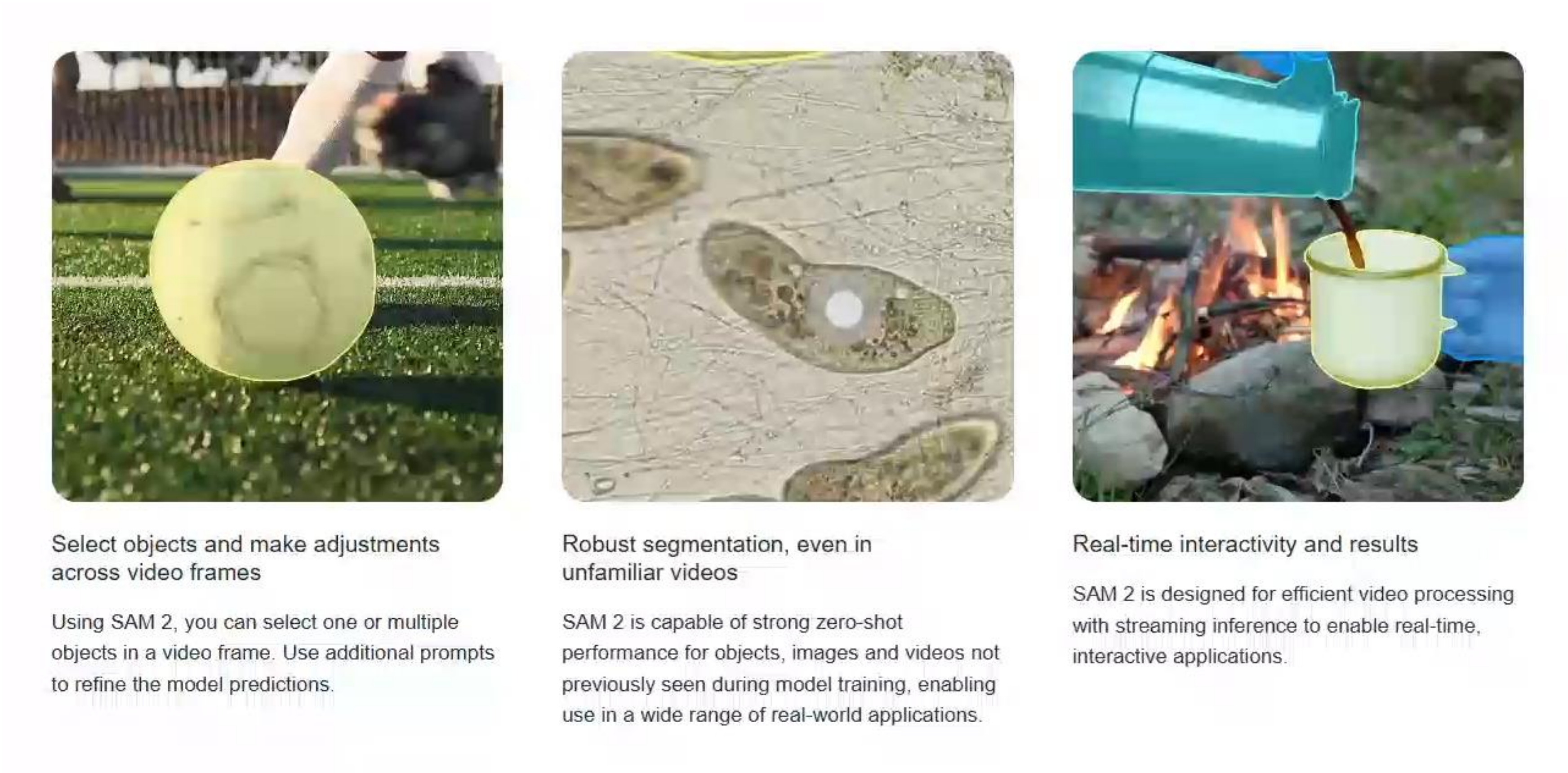
Enter your prompt

Here's a CSV. Can you describe what's in it?

Meta AI – LLaMA 3.1

https://ai.meta.com/blog/meta-llama-3-1/

Era of Generative AI (GenAI) - CV



Select objects and make adjustments across video frames

Using SAM 2, you can select one or multiple objects in a video frame. Use additional prompts to refine the model predictions.

Robust segmentation, even in unfamiliar videos

SAM 2 is capable of strong zero-shot performance for objects, images and videos not previously seen during model training, enabling use in a wide range of real-world applications.

Real-time interactivity and results

SAM 2 is designed for efficient video processing with streaming inference to enable real-time, interactive applications.

Meta AI – SAM 2

https://ai.meta.com/sam2/

Era of Generative AI (GenAI) - Multimodal

llama.com - 要退出全屏，请按 Esc

LLAMA 3.2
Multimodal
0:00 / 0:38

Meta AI – LLaMA 3.2 Vision

https://ai.meta.com/blog/llama-3-2-connect-2024-vision-edge-mobile-devices/

Era of Generative AI (GenAI) - Robotics

With Spatial Intelligence, AI Will Understand the Real World | Fei-Fei Li | TED

Seeing is
for doing

With Spatial Intelligence, AI Will Understand the Real World | Fei-Fei Li | TED

https://www.youtube.com/watch?v=y8NtMZ7VGmU

Era of Generative AI (GenAI) - AI4Science

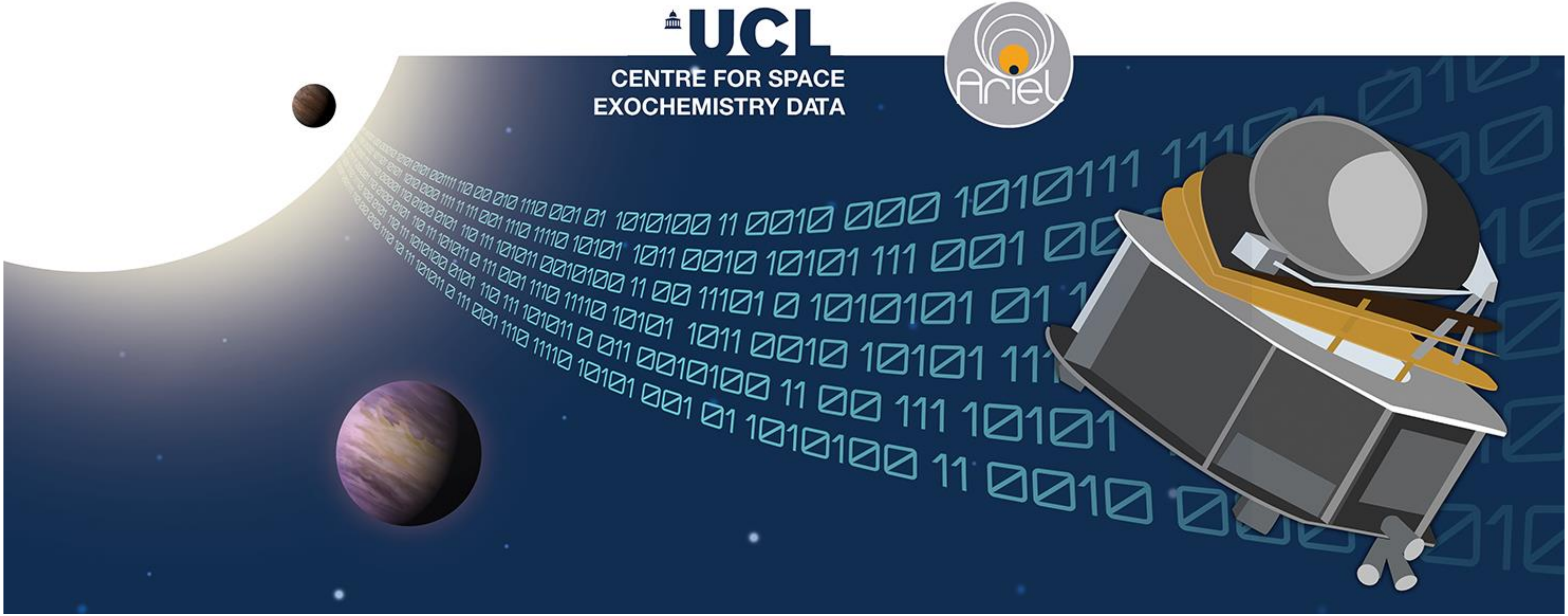

UCL
CENTRE FOR SPACE
EXOCHEMISTRY DATA
Ariel

Image Credit: NeurIPS - Ariel Data Challenge 2024 | Kaggle

References

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, *25*. https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems, 6000–6010. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020, October 2). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. International Conference on Learning Representations. International Conference on Learning Representations (ICLR). https://openreview.net/forum?id=YicbFdNTTy