



VICTORIA UNIVERSITY OF
WELLINGTON
TE HERENGA WAKA

Current state of AI

Dr. Simon McCallum

Plan



I have you for a long time

Update from last August (7 months)

Tell and Show

Added -

- Claude 3,
- Google Docs AI,
- Meta Cognitino

Update From August



Hugging face -

- Models +250K -> +400K
- Applications +100K -> +150K
- Data sets +50K -> +100K

2023

Trending on 🤖 this week

📦 Models

stabilityai/stable-diffusion-...

Updated 7 days ago • ↓ 473k • ♥ 1.66k

Qwen/Qwen-7B-Chat

Updated about 14 hour... • ↓ 31.2k • ♥ 409

meta-llama/Llama-2-7b

Updated 22 days ago • ♥ 1.68k

stabilityai/stablecode-comple...

Updated 2 days ago • ↓ 1.25k • ♥ 137

Qwen/Qwen-7B


Updated about 14 hour... • ↓ 28.5k • ♥ 183

[Browse +250k models](#)


📁 Spaces

Open LLM  Leaderboard

♥ 4.34k

MusicGen 

♥ 2.29k

Music To Image 

♥ 111

AudioLDM2 Text2Audio Text2Musi...

♥ 83

MTEB Leaderboard 

♥ 702

[Browse +100k applications](#)

📄 Datasets

fka/awesome-chatgpt-prompts

Updated Mar 7 • ↓ 3.79k • ♥ 2.93k

nampdn-ai/tiny-codes

Updated 5 days ago • ↓ 489 • ♥ 69

BAAI/COIG-PC

Updated 3 days ago • ↓ 211 • ♥ 116

Open-01ca/Open01ca

Updated 27 days ago • ↓ 10.9k • ♥ 557

timdettmers/openassistant-gua...

Updated May 27 • ↓ 38k • ♥ 154

[Browse +50k datasets](#)

Trending on 🤗 this week

📦 Models

google/gemma-7b

Updated 8 days ago • ↓ 216k • ♥ 1.99k

bigcode/starcoder2-15b

Updated about 11 hours... • ↓ 34.4k • ♥ 374

ByteDance/SDXL-Lightning

Updated 4 days ago • ↓ 355k • ♥ 1.08k

playgroundai/playground-v2.5-...

Updated 2 days ago • ↓ 51k • ♥ 325

google/gemma-7b-it

Updated 2 days ago • ↓ 131k • ♥ 879

[Browse 400k+ models](#)

🧩 Spaces

AI Comic Factory

♥ 4.28k

Playground V2.5

♥ 247

TripoSR

♥ 176

Leonardo AI Image Creator

♥ 150

Open LLM Leaderboard

♥ 8.22k

[Browse 150k+ applications](#)

📄 Datasets

microsoft/orca-math-word-prob...

Updated 2 days ago • ↓ 452 • ♥ 104

bigcode/the-stack-v2

Updated 3 days ago • ↓ 245 • ♥ 96

espnet/yodas

Updated 13 days ago • ↓ 2.61k • ♥ 63

CausalLM/Refined-Anime-Text

Updated 5 days ago • ↓ 113 • ♥ 157

argilla/dpo-mix-7k

Updated 3 days ago • ↓ 4.11k • ♥ 70

[Browse 100k+ datasets](#)

ChatGPT



Code interpreter -> Advance Data Analytics

Plugins being removed

GPTs - added and very cool context and memory

ChatGPT Team - Pay to have your Data Private

The logo for the ChatGPT Team, featuring a dark grey rectangular background. On the left, there are three blue diamonds arranged in a triangular pattern. To the right of the diamonds, the word 'Team' is written in a white, sans-serif font.

Chat Memory added

Build a custom collection of data and prompts

AI to help create - then refine

First Show session.

Liability



Pak n Save recipe (2023 funny)

Air Canada legally require to honor
Chatbot deal (2024 costly)

Copyright lawsuit vs StabilityAI mostly
dead

Bias issues



Google Gemini image generation paused because of Ethically diverse Nazi's

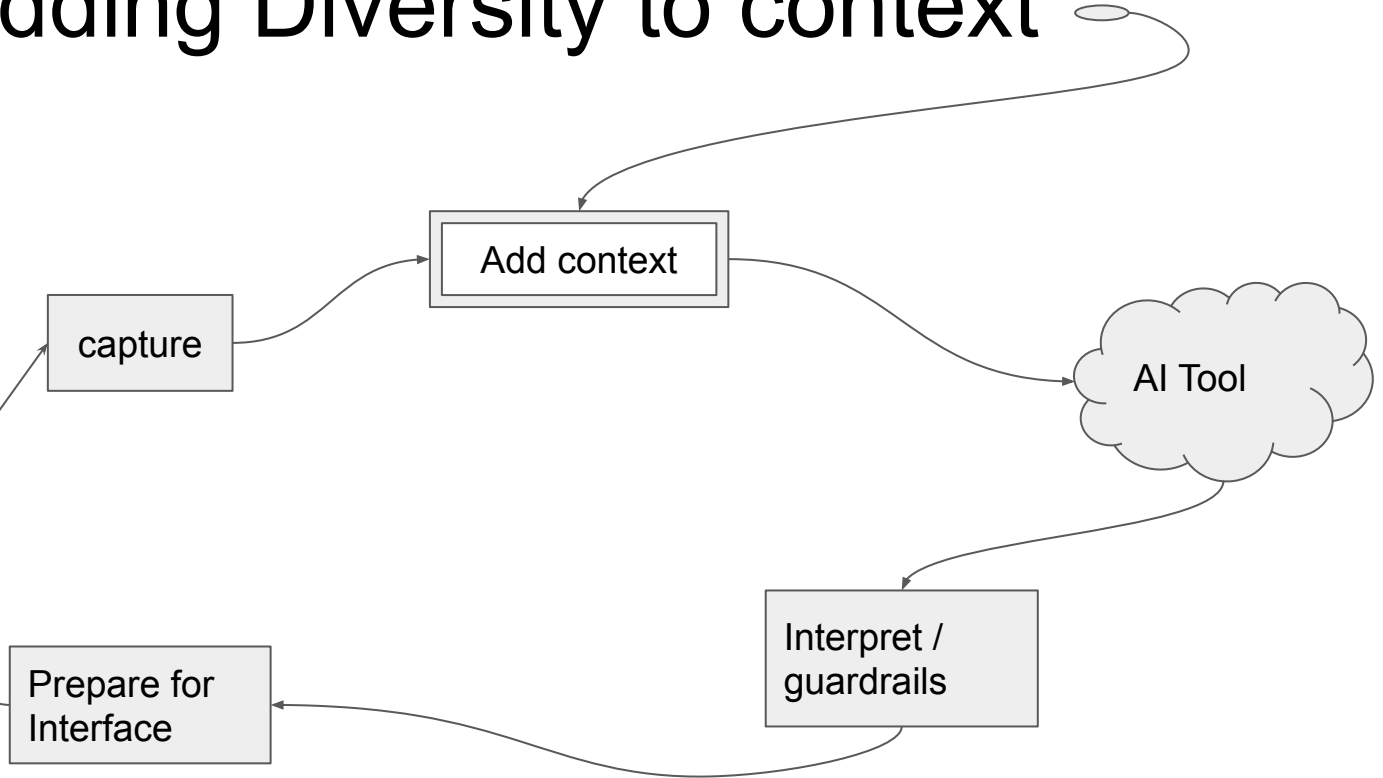
Bias is complex - adding a diversity prompt to image generation



Bias reduction



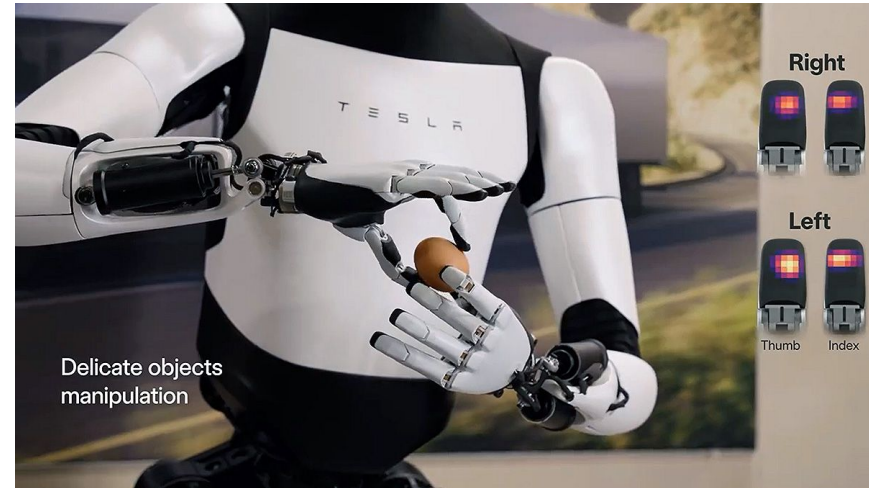
Google Adding Diversity to context



Robots



RT-2 -> X1 Elle & Optimus



Large Action Models



Rabbit AI

The idea that you train the model to do actions for you.

Video/image processing - trained on interfaces - can interact with any program.



Rabbit Process



1. Teach an interaction
2. Uses image processing to identify interface elements
3. Records interactions
4. Create Action Macros to engage in the interface.

Sora - OpenAI



Video generated from text prompt

Tyler Perry pauses

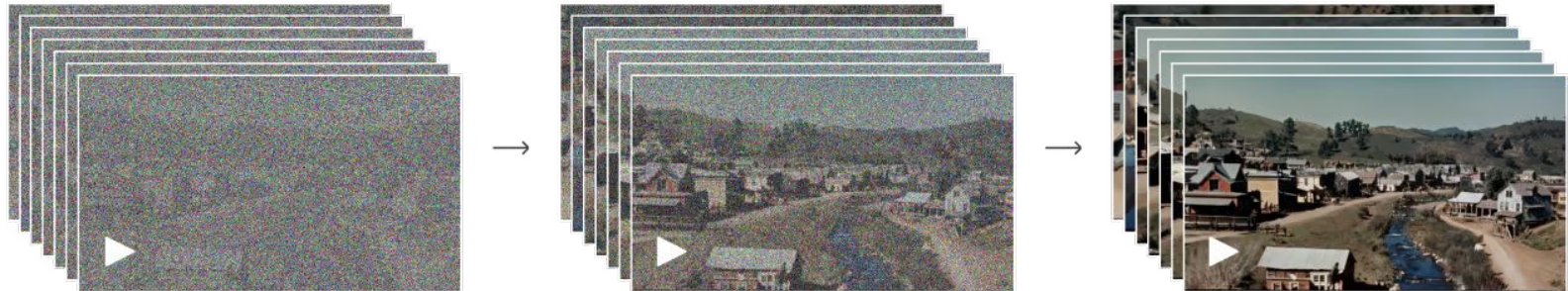
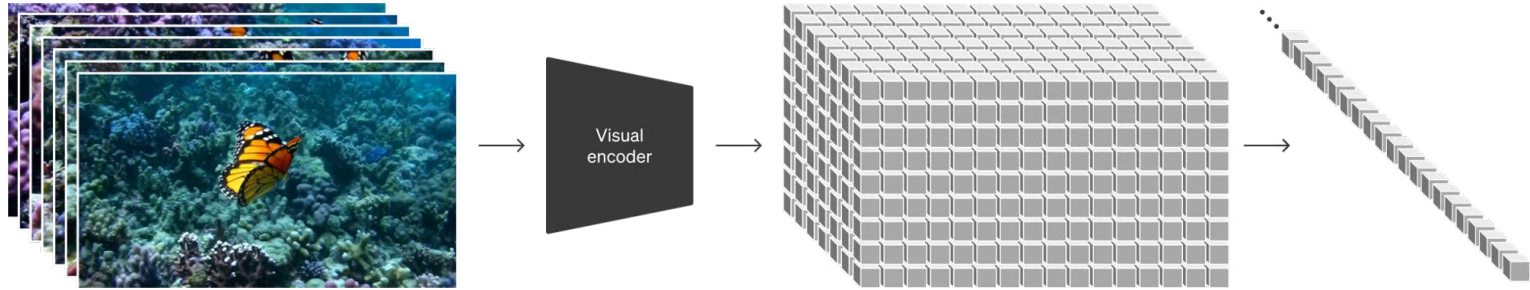
\$800 Million studio dev

Builds a 3D model of world

Improvements coming

Tactical release

Sora



Gemini



Gemini Ultra, pro, nano and Gemma

Gemini Pro 1.5 - 1 Million tokens

Gemini Ultra 32 runs - vote on output

AlphaCode 2 1 Mill - test each output,
then score, then vote



Market Summary > NVIDIA Corp

887.00 USD

+ Follow

+424.59 (91.82%) ↑ past 6 months

Closed: 6 Mar, 7:59 pm GMT-5 • Disclaimer

After hours 885.30 -1.70 (0.19%)

1D | 5D | 1M | 6M | YTD | 1Y | 5Y | Max



Open	880.22	Mkt cap	2.22T	52-wk high	897.24
High	897.24	P/E ratio	74.33	52-wk low	222.97
Low	870.30	Div yield	0.018%		

Agency



AutoGPT -> [Autogen Studio](#)

Using AI to make plans

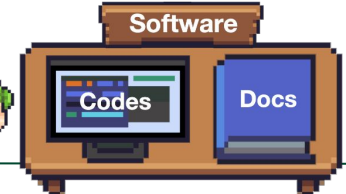
Other AIs to execute plans

The plans are getting better!

Develop a
Gomoku game



ChatDev



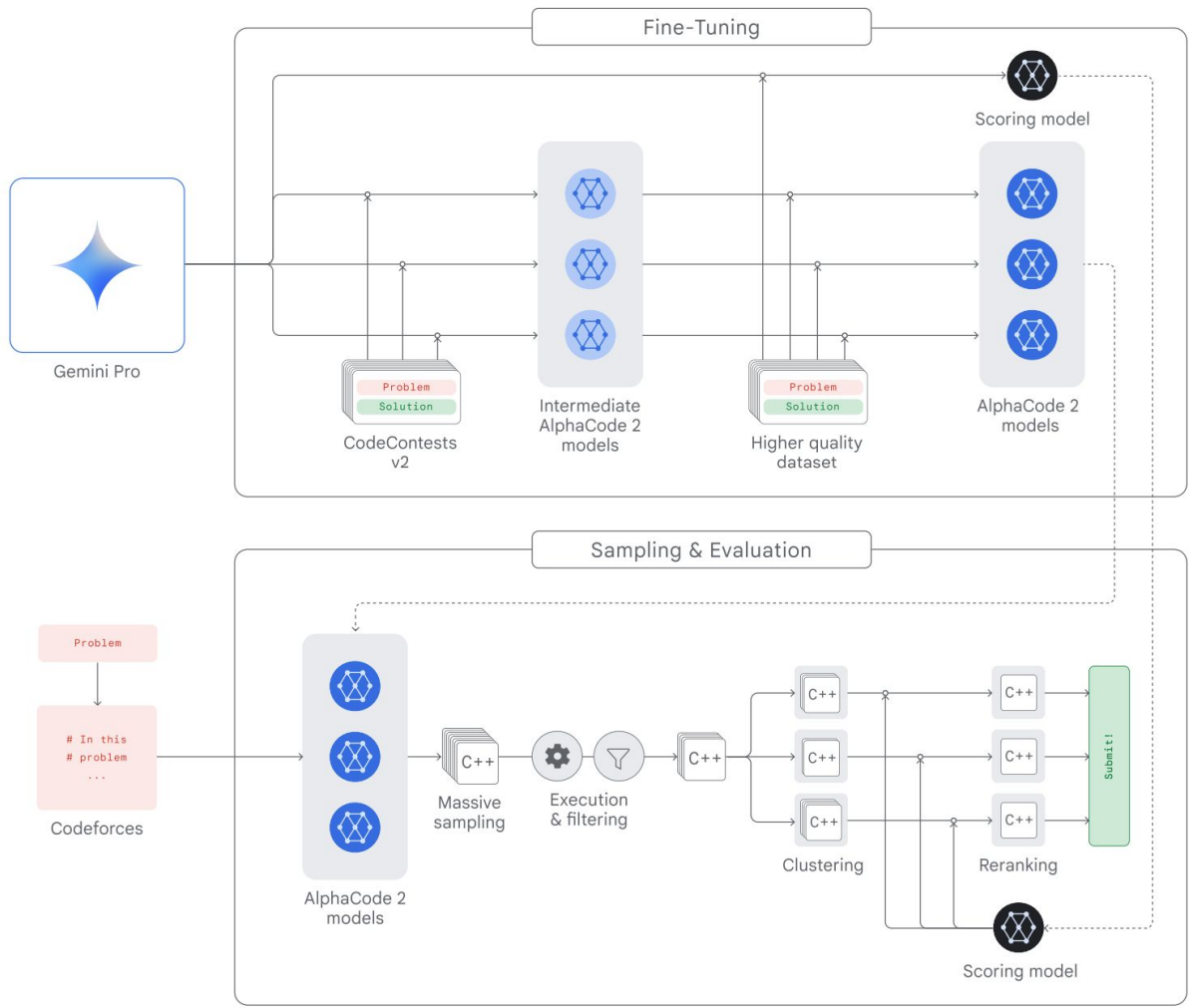
Programming
teams

Define roles

Programming



- StackOverflow - adding AI programming tools
- NVidia CEO - no programming in 5 years
- Alpha Code 2 1 Millions runs
- Claude 3 - code Human eval 83%
- Double Coding assistant vs CoPilot



Claude 3

Released 5/03/24

Best in class

Note Sonnet is better than Paid for GPT-4

OPUS \$20 US/month

	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku	GPT-4	GPT-3.5	Gemini 1.0 Ultra	Gemini 1.0 Pro
Undergraduate level knowledge <i>MMLU</i>	86.8% 5 shot	79.0% 5-shot	75.2% 5-shot	86.4% 5-shot	70.0% 5-shot	83.7% 5-shot	71.8% 5-shot
Graduate level reasoning <i>GPQA, Diamond</i>	50.4% 0-shot CoT	40.4% 0-shot CoT	33.3% 0-shot CoT	35.7% 0-shot CoT	28.1% 0-shot CoT	—	—
Grade school math <i>GSM8K</i>	95.0% 0-shot CoT	92.3% 0-shot CoT	88.9% 0-shot CoT	92.0% 5-shot CoT	57.1% 5-shot	94.4% Maj1@32	86.5% Maj1@32
Math problem-solving <i>MATH</i>	60.1% 0-shot CoT	43.1% 0-shot CoT	38.9% 0-shot CoT	52.9% 4-shot	34.1% 4-shot	53.2% 4-shot	32.6% 4-shot
Multilingual math <i>MGSM</i>	90.7% 0-shot	83.5% 0-shot	75.1% 0-shot	74.5% 8-shot	—	79.0% 8-shot	63.5% 8-shot
Code <i>HumanEval</i>	84.9% 0-shot	73.0% 0-shot	75.9% 0-shot	67.0% 0-shot	48.1% 0-shot	74.4% 0-shot	67.7% 0-shot
Reasoning over text <i>DROP, F1 score</i>	83.1 3-shot	78.9 3-shot	78.4 3-shot	80.9 3-shot	64.1 3-shot	82.4 Variable shots	74.1 Variable shots
Mixed evaluations <i>BIG-Bench-Hard</i>	86.8% 3-shot CoT	82.9% 3-shot CoT	73.7% 3-shot CoT	83.1% 3-shot CoT	66.6% 3-shot CoT	83.6% 3-shot CoT	75.0% 3-shot CoT
Knowledge Q&A <i>ARC-Challenge</i>	96.4% 25-shot	93.2% 25-shot	89.2% 25-shot	96.3% 25-shot	85.2% 25-shot	—	—
Common Knowledge <i>HellaSwag</i>	95.4% 10-shot	89.0% 10-shot	85.9% 10-shot	95.3% 10-shot	85.5% 10-shot	87.8% 10-shot	84.7% 10-shot

Graduate Question



GPQA

Graduates struggle to answer

Organic Chemistry

Methylcyclopentadiene (which exists as a fluxional mixture of isomers) was allowed to react with methyl isoamyl ketone and a catalytic amount of pyrrolidine. A bright yellow, cross-conjugated polyalkenyl hydrocarbon product formed (as a mixture of isomers), with water as a side product. These products are derivatives of fulvene. This product was then allowed to react with ethyl acrylate in a 1:1 ratio. Upon completion of the reaction, the bright yellow color had disappeared. How many chemically distinct isomers make up the final product (not counting stereoisomers)?

- A) 2
- B) 16
- C) 8
- D) 4

Genetics

If a sperm from species A is injected into an egg from species B and both species have the same number of chromosomes, what would be the main cause of the resulting zygote mortality?

- A) Species specific zona pellucida proteins on the egg cannot bind sperms from a different species.
- B) Epistatic interactions between the genes of different species
- C) Chromosomal incompatibilities will cause failure of meiosis leading to death of zygote.
- D) Chromosomal recombination will not occur in different species.

Molecular Biology

A scientist studies the stress response of barley to increased temperatures and finds a protein which contributes to heat tolerance through the stabilisation of cell membrane. The scientist is very happy and wants to create a heat-tolerant cultivar of diploid wheat. Using databases, they find a heat tolerance protein homologue and start analysing its accumulation under heat stress. Soon enough, the scientist discovers this protein is not synthesised in the wheat cultivar they study. There are many possible reasons for such behaviour, including:

- A) A miRNA targets the protein, which makes exonucleases cut it immediately after the end of translation and before processing in ER
- B) Trimethylation of lysine of H3 histone in position 27 at the promoter of the gene encoding the target protein
- C) A stop-codon occurs in the 5'-UTR region of the gene encoding the target protein
- D) The proteolysis process disrupts a quaternary structure of the protein, preserving only a tertiary structure

Context length



Google Gemini Pro 1.5 context
1 Million tokens - 700,000 words.

Claude 3 - 200,000 tokens (1 Million
possible for select customers)

These are enormous - integrated RAG

Current models



Google	<u>RT-2</u> (embodied learning)
Google	<u>Gemini Ultra</u> (Jan 20) & nano
Google	<u>AlphaGeometry</u> (85%+ Math)
Google	<u>AlphaCode 2</u>
Meta	<u>LLama 2</u> , + <u>Code LLama</u>
Open AI	<u>ChatGTP 4</u> <u>Vision</u> <u>Dall-E3</u>
Open AI	<u>Sora</u>
MS	<u>Copilot</u> & <u>AutoGen</u> <u>LlaVa</u>
Anthropic	<u>Claude 3</u> (NZ added 5 March)



AI Safety



EU Regulations last weeks presentation

This week - meta-cognition concern!

Policy tracking page in NZ

<https://www.brainbox.institute/nz-ai-tracker>

Metacognition?



Claude 3 - metacognition?

Developer at Anthropic - Claude 3 commented that the ‘needle in haystack” task appeared to be a test. - Thus identifying that the query was a test of performance.

Knowing about being tested

Conscious?

Criteria

Computation

Mortal computing

<https://arxiv.org/abs/2308.08708>

Recurrent processing theory

RPT-1: Input modules using algorithmic recurrence

RPT-2: Input modules generating organised, integrated perceptual representations

Global workspace theory

GWT-1: Multiple specialised systems capable of operating in parallel (modules)

GWT-2: Limited capacity workspace, entailing a bottleneck in information flow and a selective attention mechanism

GWT-3: Global broadcast: availability of information in the workspace to all modules

GWT-4: State-dependent attention, giving rise to the capacity to use the workspace to query modules in succession to perform complex tasks

Computational higher-order theories

HOT-1: Generative, top-down or noisy perception modules

HOT-2: Metacognitive monitoring distinguishing reliable perceptual representations from noise

HOT-3: Agency guided by a general belief-formation and action selection system, and a strong disposition to update beliefs in accordance with the outputs of metacognitive monitoring

HOT-4: Sparse and smooth coding generating a “quality space”

Attention schema theory

AST-1: A predictive model representing and enabling control over the current state of attention

Predictive processing

PP-1: Input modules using predictive coding

Agency and embodiment

AE-1: Agency: Learning from feedback and selecting outputs so as to pursue goals, especially where this involves flexible responsiveness to competing goals

AE-2: Embodiment: Modeling output-input contingencies, including some systematic effects, and using this model in perception or control

Making AI laugh



Game Jam

Send image ask AI if it is funny

Testing the emotion processing

Confidence in AI



- High degree of overconfidence
- Prompting Chain of thought improves alignment of confidence
- Consistency-based testing - repeated questions. Google Gemini Ultra 32 Samples
- Confidence in answer still challenging

Cognitive Nomad



Building cognitive castles

Moving House

Nomadic life

Rebuild vs Move on

