

## Hallucinations are (almost) all you Need

This rapid artistic overview of key scientific AI examples that covers a year (loosely defined as starting with GPT-4 on March 14th, 2023) is framed by the hypothesis that fundamental research in science is being transformed by a practice predominantly associated with the arts: namely hallucinations.

#### **JHAVE @ GLIA.CA**

CENTRE FOR DIGITAL NARRATIVE UNIVERSITY OF BERGEN



Hallucinations: neurological data which arises without a direct relationship to external stimuli.

# Simulations: hallucinations which seem





## SIMULATIONS POPULATE TOPOLOGIES OF THE POSSIBLE.

SIMULATIONS ARE A SUBSET OF HALLUCINATIONS THAT POPULATE TOPOLOGIES OF THE POSSIBLE. THEREFORE, HALLUCINATIONS ARE OFTEN INFORMATIVE, HELPFUL.

MOREOVER, IT CAN BE CLAIMED THAT SIMULATIONS (I.E. HALLUCINATIONS) ARE UBIQUITOUS, UNIVERSAL NECESSARY FEATURES OF ALL COGNITIVE SYSTEMS. ENTITIES MAKE GUESSES WITHIN A STATE SPACE OF POTENTIAL TEMPORALITIES, - PREDICTIVE SIMULATIONS THAT GENERATE PLAUSIBLE FUTURE SCENARIOS.

WHAT WILL HAPPEN NEXT? WHAT COULD BE BEST? WHAT SHOULD 'I' DO?

TO A LARGE DEGREE, THOUGHT IS THINKING ABOUT WHAT IS POSSIBLE, THEREFORE THOUGHT IS A KIND OF HALLUCINATION.

FROM THIS VIEWPOINT, SIMULATIONS AND HALLUCINATIONS ARE ROOTED IN EVOLUTIONARY NEURO-PHYSIOGNOMY. WITHOUT HALLUCINATIONS ENTITIES WOULDN'T BE ABLE TO CALCULATE TRAJECTORIES, ANTICIPATE ACTIONS, NAVIGATE SOCIAL SITUATIONS, HUNT, FEED, BREED. THE UNIVERSE WOULD BE STRANDED IN A PERPETUAL NOW.



## HALLUCINATIONS ALONE ARE NOT ENOUGH.

HALLUCINATIONS REQUIRE FACT-CHECKING BEFORE BECOMING USEFUL.

HALLUCINATIONS BECOME USEFUL WHEN FILTERED TOWARD A RIGOROUS OPTIMIZATION-TARGET GROUNDED IN SENSORIALLY-RELEVANT LOCAL CONDITIONS.

ALL FUNCTIONAL ENTITIES COMBINE HALLUCINATIONS WITH ASSESSMENT-FILTRATION SYSTEMS. HALLUCINATIONS THAT SURVIVE THE FACT-FILTER BECOME KNOWN AS SIMULATIONS.

OFTEN SIMULATIONS ARISE RAPIDLY AND INSTINCTIVELY. ASSESSMENT OCCURS SLOWLY AND METICULOUSLY. IF HALLUCINATIONS OCCUR IN THE ABSENCE OF ROBUST ACCURATE CREDIBLE ASSESSMENT, THEN MENTAL ILLNESS AND COGNITIVE DYSPHORIA EMERGES.



#### PERCEPTIONS AND MEMORIES ARE ALSO SIMULATIONS (AND THUS PARTIALLY HALLUCINATIONS)

THE WORLD'S APPEARANCE, ITS PHENOMENA (TREES, BEINGS, CULTURE, CELLPHONES, AND RESEARCH REPORTS...) ALL EMERGE IN A CATACLYSMIC, CHURNING, TENDER PHENOMENOLOGICAL FRENZY OF DATA DESIGNED TO ENTER ANY SENSOR WITH CAPACITY FOR THAT RANGE OF REVERBERATION (VISION, SCENT, TOUCH, TASTE OR SOUND).

AS NEUROPHYSIOLOGY ABSORBS PHENOMENA, IT DOWNSAMPLES (REDUCES THE RESOLUTION OF) A SEEMINGLY EXTERIOR WORLD —AND THEN BUILDS A SIMULATION, A MEMORY, A STORAGE-TRACE WITHIN ITS OWN BODY-MIND. THE BODY-MIND BECOMES A REPOSITORY FOR THE PAST (A SIMULATION OF WHAT IT HAS SEEN), AN ENACTMENT OF THE FUTURE (A SIMULATION OF WHAT IT EXPECTS), AND CREATIVITY (AN IMAGINATION OF MEANING, STRUCTURE, ENQUIRY INTO WHAT IS OR WHAT COULD BE).

## INTROCEPTION

**NEUROLOGISTS NOW SPEAK OF INTERNAL PERCEPTIONS AS INTEROCEPTION.** VISCERA OF THE BODY GENERATES WAVE AFTER WAVE OF DATA THAT INUNDATE THE INSULA AND CINGULATE REGIONS.

SO IT'S NOT JUST EXTERIORITY WHICH GIVES RISE TO A SENSE OF A WORLD, BUT THE INTERIORITY OF THE BODY'S CHAMBERS, TUBES, ORGANS AND DISTRIBUTED TRIBUTARIES OF SENSORIAL FEEDBACK WHICH CONSTRUCT A HOMEOSTATIC **MODEL OF WHAT IDENTITY IS.** 

HOW BODY FEELS BECOME WHAT IDENTITY FEELS THAT IT IS FEELING.

THE INNER SELF-MODEL OF THE MIND IS TO SOME DEGREE ALSO A SIMULATION

#### "NOBODY EVER WAS OR HAD A SELF.

ALL THAT EVER EXISTED WERE CONSCIOUS SELF-MODELS THAT COULD NOT BE RECOGNIZED AS MODELS. THE PHENOMENAL SELF IS NOT A THING, BUT A PROCESS — AND THE SUBJECTIVE EXPERIENCE OF BEING SOMEONE EMERGES IF A CONSCIOUS INFORMATION-PROCESSING SYSTEM OPERATES UNDER A TRANSPARENT SELF-MODEL. YOU ARE SUCH A SYSTEM RIGHT NOW, AS YOU READ THESE SENTENCES. "

Being No One, Thomas Metzinger (2003)

#### TECHNOLOGICAL **APPROACH TO MIND EVERYWHERE** (TAME): **AN EXPERIMENTALLY-GROUNDED FRAMEWORK FOR UNDERSTANDING DIVERSE BODIES AND MINDS**

**"TAME - TECHNOLOGICAL APPROACH TO MIND EVERYWHERE - A FRAMEWORK FOR UNDERSTANDING AND MANIPULATING COGNITION IN UNCONVENTIONAL SUBSTRATES. TAME FORMALIZES A NON-BINARY (CONTINUOUS), EMPIRICALLY-BASED APPROACH TO STRONGLY EMBODIED AGENCY. WHEN APPLIED TO REGENERATING/DEVELOPMENTAL SYSTEMS, TAME SUGGESTS A PERSPECTIVE ON MORPHOGENESIS AS AN EXAMPLE OF BASAL COGNITION. ... THIS ARCHITECTURE OF MULTI-SCALE COMPETENCY OF BIOLOGICAL SYSTEMS HAS IMPORTANT IMPLICATIONS FOR PLASTICITY OF BODIES AND MINDS, GREATLY POTENTIATING EVOLVABILITY."** 

MICHAEL LEVIN, 24 DECEMBER 2021



### HALLUCINATIONS

HALLUCINATIONS IN PEOPLE ARE **CONVENTIONALLY ASSOCIATED WITH** MENTAL ILLNESS, DRUGS, AND/OR GENIUS.

HALLUCINATIONS IN AI (MOSTLY IN LARGE LANGUAGE MODELS) HAVE BEEN CRITIQUED **AS NET-NEGATIVES: CONTRIBUTING TO DISINFORMATION, BIAS, POST-TRUTH, DEEP-**FAKES, COLLAPSE OF DEMOCRACY, **COPYRIGHT THEFT, ETC...** 



### YET, THE BENEFIT OF AI HALLUCINATIONS

OF PROTEINS/CRYSTALS/ ALGORITHMS/CIRCUITS ETC ... PRUNED DOWN TO THE FEASIBLE, ARE CONTRIBUTING TO A REVOLUTIONARY ACCELERATION OF GROUNDBREAKING SCIENTIFIC DISCOVERIES.



#### HALLUCINATIONS ARE FOUNDATIONAL

TO ART, CULTURE, SCIENCE, SPIRITUALITY, LITERATURE

WITHOUT VISIONARY THINKERS, LANGUAGE, MATHEMATICS, PHYSICS AND THE ARTS WOULDN'T EVEN EXIST.

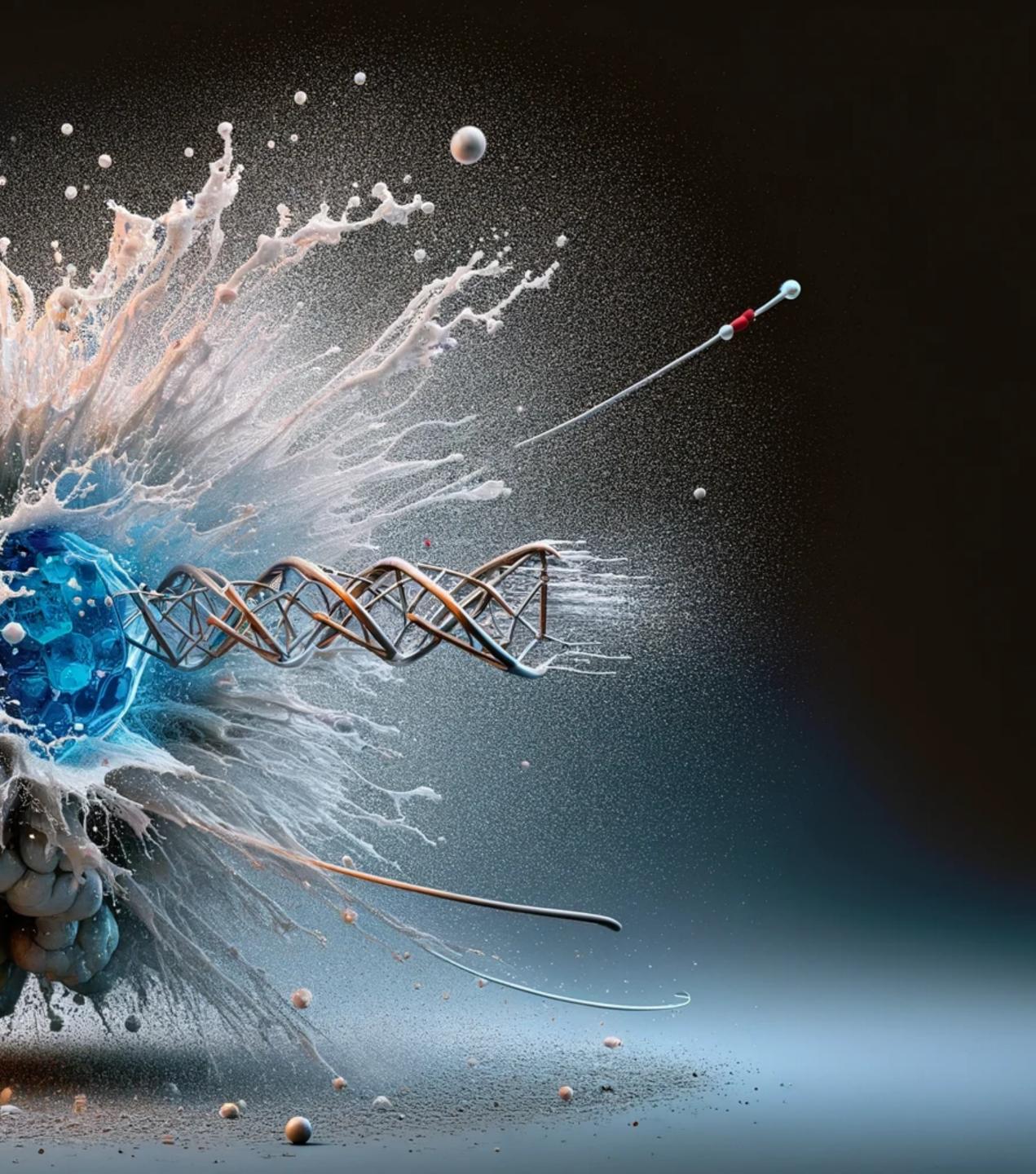


A LONG, BOUNDLESS, AND SYSTEMATIZED DISORGANIZATION OF ALL THE SENSES. **ARTHUR RIMBAUD, "LETTRES DU VOYANT" (LETTERS OF THE SEER). 1871.** 

# A POET MAKES HIMSELF A VISIONARY THROUGH

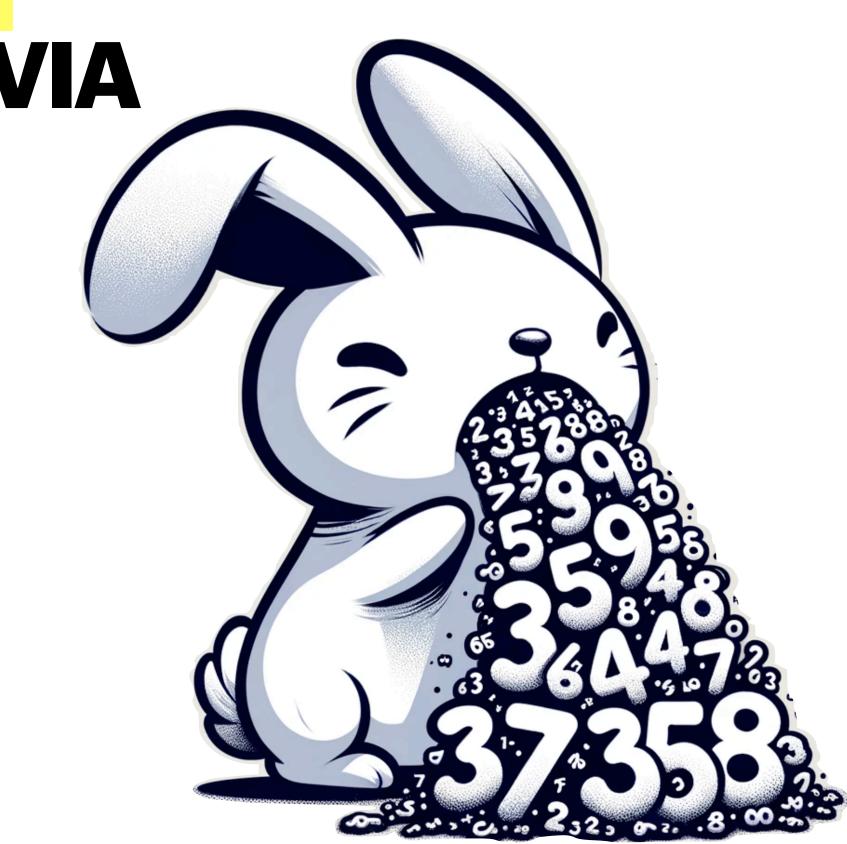
## HALLUCINATIONS ARE FUNDAMENTAL

TO RESEARCH SCIENCE AS AN ART IN THE AGE OF INCIPIENT A.G.I.



## A MAJOR LIMITATION OF AUTOMATED THEOREM PROVERS COMPARED TO HUMANS -- THE GENERATION OF ORIGINAL MATHEMATICAL TERMS --MIGHT BE ADDRESSABLE VIA GENERATION FROM LANGUAGE MODELS.

STANISLAS POLU, ILYA SUTSKEVER, GENERATIVE LANGUAGE MODELING FOR AUTOMATED THEOREM PROVING SEPTEMBER 7, 2020



Hallucinations underlie contemporary research into #GenAl, numeric and algorithmic optimizations, AI designing hardware accelerators. reward mechanisms, non-invasive brain sensors, protein folding, de novo drug discovery, sustainable deep-tech materials, autonomous gamified robotic chemistry labs, AI solving logic and geometry puzzles, and neuromorphic organoid computing.

#GENAI OPTIMIZATIONS ACCELERATORS TELEPATHY DRUG DISCOVERY DEEP TECH AUTONOMOUS LABS ORGANOID COMPUTING MATH & REASONING



## 

#### Hallucinations all the way down



# #GENAI

GPT-4, Dalle.3, MidJourney, Stable Diffusion, Suno, Pika, Runway, Genie, Magnific, Falcon, Llama2, Sora…







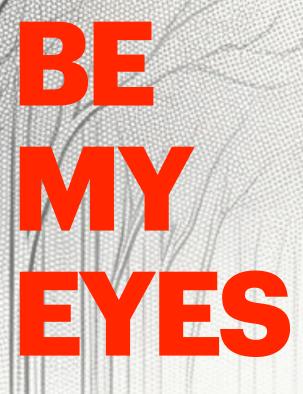
#### "while less capable than humans in many real-world scenarios, exhibits human-level performance on various professional and academic benchmarks... it passes a simulated bar exam with a score around the top 10% of test takers" OpenAI. 14 March 2023

1111

- rere

**GPT-4** 





"Bringing sight to blind and low-vision people. ~ AI-powered visual assistance for instantaneous image-to-text generation.... instantaneous identification, interpretation and conversational visual assistance for a wide variety of tasks." OpenAI GPT-4 Launch Partner. 14 March 2023



## LLAMA 2: OPEN FOUNDATION AND FINE-TUNED CHAT MODELS

"Llama 2 pretrained models are trained on 2 trillion tokens, and have double the context length than Llama 1. Its fine-tuned models have been trained on over 1 million human annotations." Meta. 18 July 2023



# DALLE.3

## **IMPROVING IMAGE GENERATION WITH BETTER CAPTIONS**

WE SHOW THAT PROMPT FOLLOWING ABILITIES OF TEXT-TO-IMAGE MODELS CAN BE SUB-STANTIALLY IMPROVED BY TRAINING ON HIGHLY DESCRIPTIVE GENERATED IMAGE CAPTIONS. EXISTING TEXT-TO-IMAGE MODELS STRUGGLE TO FOLLOW DETAILED IMAGE DESCRIPTIONS AND OFTEN IGNORE WORDS OR CONFUSE THE MEANING OF PROMPTS. WE HYPOTHESIZE THAT THIS ISSUE STEMS FROM NOISY AND INACCURATE IMAGE CAPTIONS IN THE TRAINING DATASET. WE ADDRESS THIS BY **TRAINING A BESPOKE IMAGE CAPTIONER AND USE IT TO RECAPTION THE TRAINING DATASET. WE THEN TRAIN SEVERAL TEXT-TO-IMAGE MODELS AND FIND THAT TRAINING ON THESE SYNTHETIC CAPTIONS RELIABLY IMPROVES PROMPT FOLLOWING ABILITY** FINALLY, WE USE THESE FINDINGS TO BUILD DALL-E 3: A NEW TEXT-TO-IMAGE GENERATION SYSTEM, AND BENCHMARK ITS PERFORMANCE ON AN EVALUATION DESIGNED TO MEASURE PROMPT FOLLOWING, COHERENCE, AND AESTHETICS, FINDING THAT IT COMPARES FAVORABLY TO COMPETITORS. WE PUBLISH SAMPLES AND CODE FOR THESE EVALUATIONS SO THAT FUTURE RESEARCH CAN CONTINUE OPTIMIZING THIS IMPORTANT ASPECT OF TEXT-TO-IMAGE SYSTEMS.

**OpenAI**. 20th September 2023 <u>https://openai.com/dall-e-3</u>

### EMU VIDEO: FACTORIZING TEXT-TO-VIDEO GENERATION BY EXPLICIT IMAGE CONDITIONING

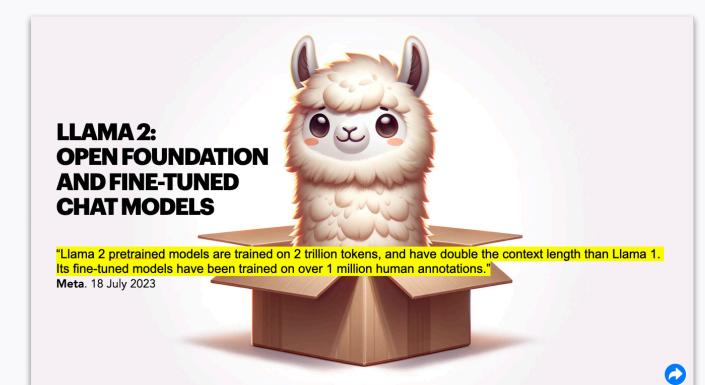
"In human evaluations, our generated videos are strongly preferred in quality compared to all prior work--81% vs. Google's Imagen Video, 90% vs. Nvidia's PYOCO, and 96% vs. Meta's Make-A-Video. Our model outperforms commercial solutions such as RunwayML's Gen2 and Pika Labs... our generations are preferred 96% over prior work." Meta. 17 November 2023

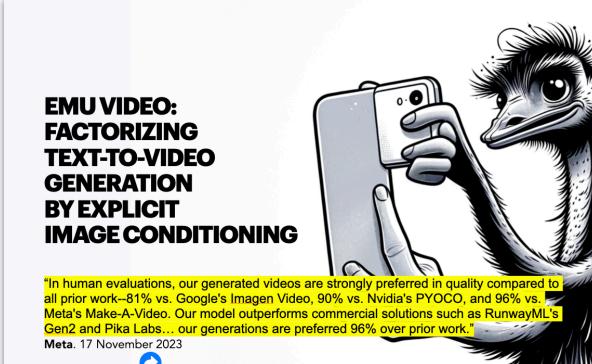


### MAMBA: LINEAR-TIME SEQUENCE MODELING WITH SELECTIVE STATE SPACES

"Mamba enjoys fast inference (5x higher throughput than Transformers) and linear scaling in sequence length, and its performance improves on real data up to million-length sequences. As a general sequence model backbone, Mamba achieves state-of-the-art performance across several modalities such as language, audio, and genomics. On language modeling, our Mamba-3B model outperforms Transformers of the same size and matches Transformers twice its size, both in pretraining and downstream evaluation." Gu & Dao. 1 December 2023









### SEMIOTIC NOTE ANTHROPOMORPHIC NAMING CONVENTIONS: EMU, LLAMA, MAMBA.

Incipient AGI, proto-organisms, implicit agency invoked by acronym, eco-nomenclature, archetypal eden projection, 6th extinction guilt replacing animals with algorithms.

Also note: the implicit sympathy requested by the Aigenerated cuteness meme critters and goat-kitty rainbows of these slides.



Google DeepMind. December 06, 2023



# SORA

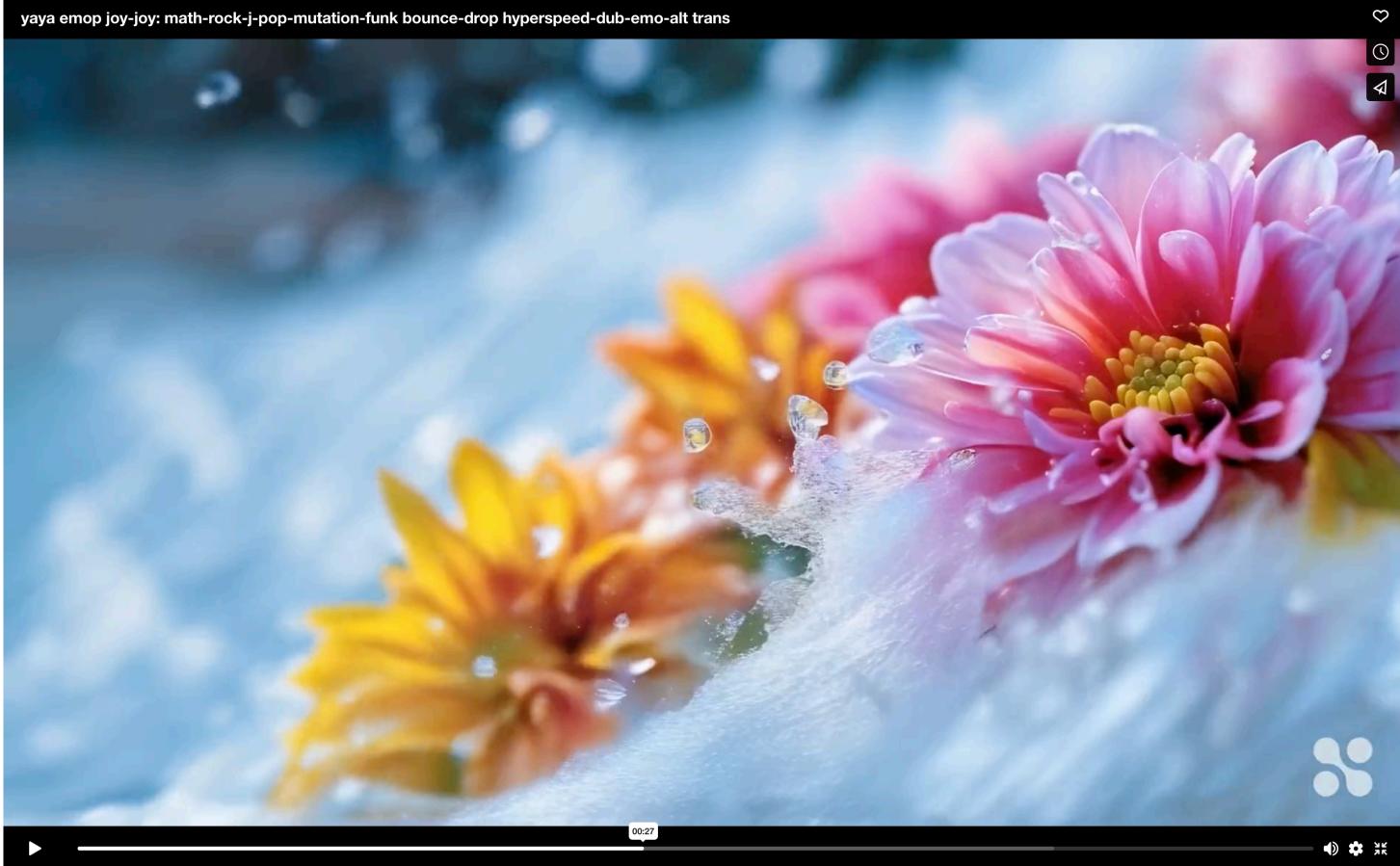
#### VIDEO GENERATION MODELS AS WORLD SIMULATORS

SORA IS ABLE TO GENERATE COMPLEX SCENES WITH MULTIPLE CHARACTERS, SPECIFIC TYPES OF MOTION, AND ACCURATE DETAILS OF THE SUBJECT AND BACKGROUND. THE MODEL UNDERSTANDS NOT ONLY WHAT THE USER HAS ASKED FOR IN THE PROMPT, BUT ALSO HOW THOSE THINGS EXIST IN THE PHYSICAL WORLD.

**OpenAI**, Feb 15th, 2024



#### Emop: on vimeo [a one-day ai-music-video]





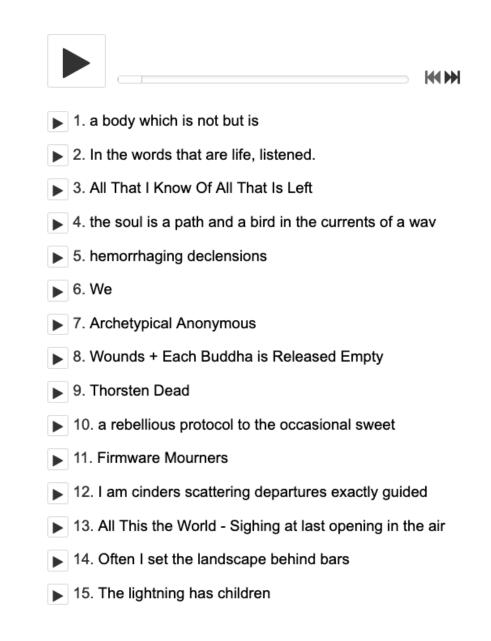






#### Suno v3 (Alpha) w. ReRites Lyrics

All songs prompted on March 6, 2024



# -> #GENALART CONCLUSIONS

GENERATIVE ALIS, OBVIOUSLY, HALLUCINATIONS.

THE VALIDITY AND **FILTRATION** OF THOSE HALLUCINATIONS ARE BASED UPON PEOPLE'S **AESTHETIC JUDGMENT OR FACTUAL CRITERIA**.

IS IT RELEVANT TO THE **INTERNAL FELT SENSE** OF A LIVED PHENOMENOLOGICAL WORLD, AN EMOTIONAL REALITY, OR A PSYCHOLOGICAL INTEGRITY?

THERE IS NO SPECIFIC USE CASE EXCEPT FOR THE **HEDONIC FEEDBACK** CIRCUIT OF OSMOTIC PLEASURE THAT ARISES WHEN A BODY ENCOUNTERS VARIATIONS, WHICH EXTEND OR REINFORCE THE REWARD MECHANISMS OF WHAT IS COMMONLY CALLED BEAUTY OR TRUTH.

6.0



# 

#### Hallucinated matrix-math, preferences, prompts, benchmarks, corpuses.



## CHINCHILLA: TRAINING COMPUTE-OPTIMAL LARGE LANGUAGE MODELS

"for every doubling of model size the number of training tokens should also be doubled. We test this hypothesis by training a predicted compute-optimal model, Chinchilla, that uses the same compute budget as Gopher but with 70B parameters and 4x more more data. Chinchilla uniformly and significantly outperforms Gopher (280B), GPT-3 (175B), Jurassic-1 (178B), and Megatron-Turing NLG (530B) on a large range of downstream evaluation tasks. This also means that Chinchilla uses substantially less compute for fine-tuning and inference, greatly facilitating downstream usage. As a highlight, Chinchilla reaches a state-of-the-art average accuracy of 67.5% on the MMLU benchmark, greater than a 7% improvement over Gopher.." → Al Optimizations DeepMind. 29 Mar 2022



## QLORA: EFFICIENT FINETUNING OF QUANTIZED LLMS

"an efficient finetuning approach that reduces memory usage enough to finetune a 65B parameter model on a single 48GB GPU while preserving full 16-bit finetuning task performance. QLoRA backpropagates gradients through a frozen, 4-bit quantized pretrained language model into Low Rank Adapters~(LoRA). Our best model family, which we name Guanaco, outperforms all previous openly released models on the Vicuna benchmark, reaching 99.3% of the performance level of ChatGPT while only requiring 24 hours of finetuning on a single GPU.... Our results show that QLoRA finetuning on a small high-quality dataset leads to state-of-the-art results, even when using smaller models than the previous SoTA. We provide a detailed analysis of chatbot performance based on both human and GPT-4 evaluations showing that GPT-4 evaluations are a cheap and reasonable alternative to human evaluation."

→ Al Optimizations Arxiv. 23 May, 2023





### GUANACO

#### **QLORA: Efficient Finetuning of Quantized LLMs**

**Tim Dettmers**\*

Artidoro Pagnoni\*

Ari Holtzman

#### Luke Zettlemoyer

University of Washington {dettmers, artidoro, ahai, lsz}@cs.washington.edu

#### Abstract

We present QLORA, an efficient finetuning approach that reduces memory usage enough to finetune a 65B parameter model on a single 48GB GPU while preserving full 16-bit finetuning task performance. QLORA backpropagates gradients through a frozen, 4-bit quantized pretrained language model into Low Rank Adapters (LoRA). Our best model family, which we name Guanaco, outperforms all previous openly released models on the Vicuna benchmark, reaching 99.3% of the performance level of ChatGPT while only requiring 24 hours of finetuning on a single GPU. QLORA introduces a number of innovations to save memory without sacrificing performance: (a) 4-bit NormalFloat (NF4), a new data type that is information theoretically optimal for normally distributed weights (b) Double Quantization to reduce the average memory footprint by quantizing the quantization constants, and (c) Paged Optimizers to manage memory spikes. We use QLORA to finetune more than 1,000 models, providing a detailed analysis of instruction following and chatbot performance across 8 instruction datasets, multiple model types (LLaMA, T5), and model scales that would be infeasible to run with regular finetuning (e.g. 33B and 65B parameter models). Our results show that QLoRA finetuning on a small high-quality dataset leads to state-of-the-art results, even when using smaller models than the previous SoTA. We provide a detailed analysis of chatbot performance based on both human and GPT-4 evaluations showing that GPT-4 evaluations are a cheap and reasonable alternative to human evaluation. Furthermore, we find that current chatbot benchmarks are not trustworthy to accurately evaluate the performance levels of chatbots. A lemon-picked analysis demonstrates where Guanaco fails compared to ChatGPT. We release all of our models and code,



## LET'S VERIFY STEP BY

"Our process-supervised model solves 78% of problems from a representative subset of the MATH test set. Additionally, we show that active learning significantly improves the efficacy of process supervision. To support related research, we also release PRM800K, the complete dataset of 800,000 step-level human feedback labels used to train our best reward model." Reasoning step-by-step, generating (hallucinating) a huge sample, selecting best. → Al Optimizations OpenAI. 31 May, 2023



#### DIRECT PREFERENCE OPTIMIZATION: YOUR LANGUAGE MODEL IS SECRETLY A REWARD MODEL

"DPO can fine-tune LMs to align with human preferences as well as or better than existing methods. Notably, fine-tuning with DPO exceeds PPO-based RLHF in ability to control sentiment of generations, and matches or improves response quality in summarization and single-turn dialogue while being substantially simpler to implement and train... building Al systems that are safe, performant, and controllable ... In this paper, we show how to directly optimize a language model to adhere to human preferences, without explicit reward modeling or reinforcement learning."

→ Al Optimizations Arxiv. 29 May, 2023



MMMU: A MASSIVE<br/>MULTI-DISCIPLINE<br/>MULTIMODAL<br/>UNDERSTANDING<br/>AND REASONING<br/>BENCHMARK<br/>FOR EXPERT<br/>AGI

"Unlike existing benchmarks, MMMU focuses on advanced perception and reasoning with domain-specific knowledge, challenging models to perform tasks akin to those faced by experts. ... Even the advanced GPT-4V and Gemini Ultra only achieve accuracies of 56% and 59% respectively, indicating significant room for improvement. We believe MMMU will stimulate the community to build next-generation multimodal foundation models towards expert artificial general intelligence."

20

APRE 4 3 A 8 A 4 4 4 2 7 5 5 8 7 3 1015 1,7 1,17 100 X 16 16 1 7 5 3 5 8 7 3

13

→ AI Optimizations
Arxiv. 27 November 2023

## PHI-2: THE SURPRISING POWER OF SMALL LANGUAGE MODELS

"<u>Phi-2</u> a 2.7 billion-parameter language model that demonstrates outstanding reasoning and language understanding capabilities, showcasing state-of-the-art performance among base language models with less than 13 billion parameters. On complex benchmarks Phi-2 matches or outperforms models up to 25x larger, thanks to new innovations in model scaling and training data curation." "<u>Textbooks Are All You Need.</u>" **Hallucinating from a solid foundation grounded in fact** leads to a proliferation of plausible potentialities. **Microsoft**. December 12, 2023

## 

HERE, WE WITNESS AN ECOSYSTEM OF AI RESEARCH EXTENDING ITSELF INTO MODULAR CIRCUITS OF SELF-IMPROVEMENT THAT OPTIMIZE AND STIMULATE INTERCOMMUNICATION AT DIVERSE LEVELS: NUMERIC, ALGORITHMIC, AND DATA (QUALITY AND SCALE).





# 

### Hallucinations of circuits, algorithms, code.





### A GRAPH PLACEMENT METHODOLOGY FOR FAST CHIP DESIGN

"we pose chip floorplanning as a reinforcement learning problem, and develop an edge-based graph convolutional neural network architecture capable of learning rich and transferable representations of the chip. As a result, our method utilizes past experience to become better and faster at solving new instances of the problem, allowing chip design to be performed by artificial agents with more experience than any human designer. Our method was used to design the next generation of Google's artificial intelligence (AI) accelerators, and has the potential to save thousands of hours of human effort for each new generation. Finally, we believe that more powerful AI-designed hardware will fuel advances in AI, creating a **symbiotic** relationship between the two fields"

→ AI Designing AI Google. 09 June 2021

Mirhoseini, A., Goldie, A., Yazgan, M. et al. A graph placement methodology for fast chip design. Nature 594, 207–212 (2021)



# DISCOVERING NOVEL ALGORITHMS WITH ALPHATENSOR

"we converted the problem of finding efficient algorithms for matrix multiplication into a single-player game ... to play this game well, one needs to **identify the tiniest of needles in a gigantic haystack** of possibilities ... **AlphaTensor**'s algorithm improves on Strassen's two-level algorithm [for 4 × 4 matrix multiplication] in a finite field for the first time since its discovery 50 years ago... Moreover, **AlphaTensor** also discovers a diverse set of algorithms with state-of-the-art complexity – up to thousands of matrix multiplication algorithms for each size, showing that the space of matrix multiplication algorithms is richer than previously thought. ... These algorithms multiply large matrices 10-20% faster" → Al Designing Al Designing Al



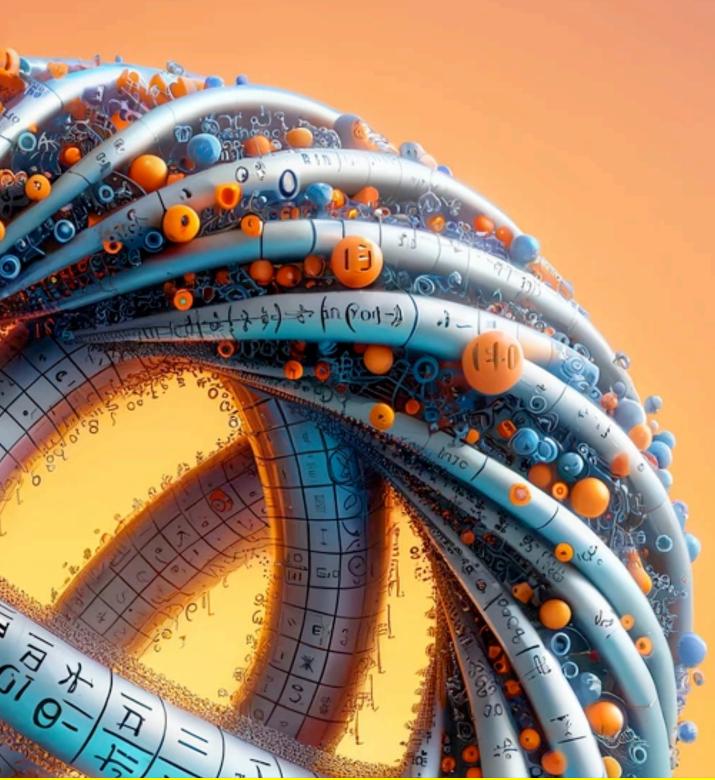
## ALPHA DEV

"Fundamental algorithms such as sorting or hashing are used trillions of times on any given day... an artificial intelligence (AI) system that uses reinforcement learning to discover enhanced computer science algorithms – surpassing those honed by scientists and engineers over decades. ... we transformed sorting into a single player 'assembly game'. ... AlphaDev uncovered new sorting algorithms that led to improvements in the LLVM libc++ sorting library that were up to 70% faster for shorter sequences ... AlphaDev's new hashing algorithm was released into the opensource <u>Abseil library</u>," → AI Designing AI Daniel J. Mankowitz and Andrea Michi. DeepMind. 07June 2023

## PROJECT CEIBA

"NVIDIA and AWS are partnering on Project Ceiba to design the world's fastest GPU-powered AI supercomputer—an at-scale system with GH200 NVL32 and Amazon EFA interconnect hosted by AWS for NVIDIA's own research and development team. This first-of-its-kind supercomputer—featuring 16,384 NVIDIA GH200 Superchips and capable of processing 65 exaflops of AI—will be used by NVIDIA to propel its next wave of generative AI innovation."

NVIDIA and AWS. 28 November 2023





# NVIDIA DGX SUPERPOD

"NVIDIA today announced its next-generation AI supercomputer — the NVIDIA DGX SuperPOD™ powered by NVIDIA GB200 Grace Blackwell Superchips — for processing trillion-parameter models with constant uptime for superscale generative AI training and inference workloads. Featuring a new, highly efficient, liquid-cooled rack-scale architecture, the new DGX SuperPOD is built with NVIDIA DGX™ GB200 systems and provides **11.5 exaflops** of AI supercomputing at FP4 precision and 240 terabytes of fast memory — scaling to more with additional racks. Each DGX GB200 system features 36 NVIDIA GB200 Superchips — which include 36 NVIDIA Grace CPUs and 72 NVIDIA Blackwell GPUs — connected as one supercomputer via fifth-generation NVIDIA NVLink®. GB200 Superchips deliver up to a 30x performance increase compared to the NVIDIA H100 Tensor Core GPU for large language model inference workloads.

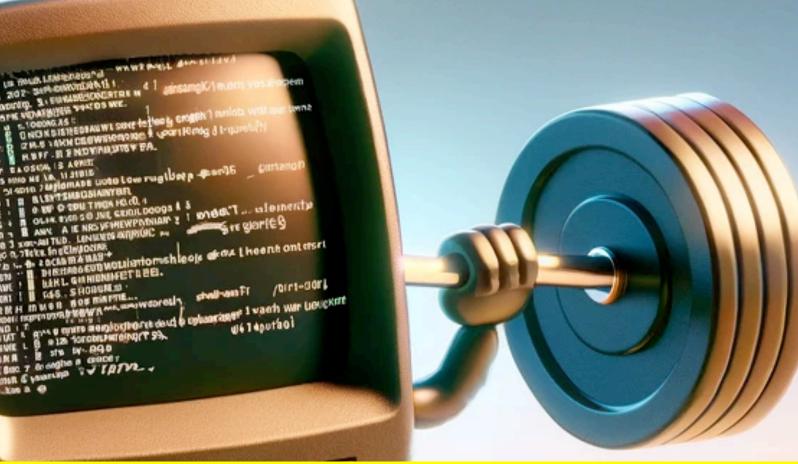
→ AI Designing AI NVIDIA. 18 March 2024





# ALPHA CODE 2

"AlphaCode 2 relies on the combination of powerful language models and a bespoke search and reranking mechanism. When evaluated on the same platform as the original AlphaCode, we found that AlphaCode 2 solved  $1.7 \times$  more problems, and performed better than 85% of competition participants... uses transformer-based language models to generate code at an unprecedented scale, and then smartly filters to a small set of promising programs"  $\rightarrow$  Al Designing Al Google DeepMind. December 06, 2023



# EVOLVING NEW FOUNDATION MODELS: UNLEASHING THE POWER OF AUTOMATING MODEL DEVELOPMENT

"Evolutionary Model Merge, a general method that uses evolutionary techniques to efficiently discover the best ways to combine different models from the vast ocean of different open-source models with diverse capabilities. As of writing, <u>Hugging Face</u> has over 500k models in dozens of different modalities that, in principle, could be combined to form new models with new capabilities! By working with the vast collective intelligence of existing open models, our method is able to *automatically* create *new* foundation models with desired capabilities specified by the user."

→ AI Designing AI
Sakana.ai 21 March 2024



# -> AI DESIGNING AI CONCLUSIONS

AT THE **MACHINE CODE** LEVEL, AI IS IMAGINING OPTIMIZED ALGORITHMS FOR DATA FLOW. AT THE **HARDWARE** LEVEL, THERE IS THE TPU-HARDWARE-ENHANCING CHIP-FLOOR DESIGN AI. AT THE **HUMAN** LEVEL, THERE IS ALPHA CODE, WHICH WRITES EXPERT HUMAN-READABLE CODE.

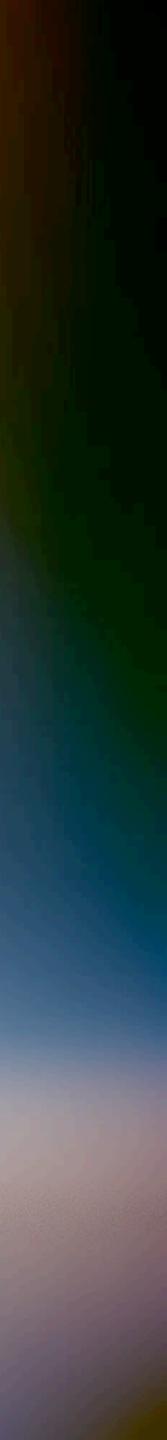
AND THEN THERE'S SAKANA, WHICH IS AN EVOLUTIONARY AI MODEL THAT DESIGNS AI MODELS.





# -> Al Mind-Reading

### Hallucinated thoughts, intentions, words, images, gestures.



### WALKING NATURALLY AFTER SPINAL CORD INJURY USING A BRAIN-SPINE INTERFACE

"digital bridge between the brain and spinal cord that enabled an individual with chronic tetraplegia to stand and walk naturally in community settings. This brain–spine interface (BSI) consists of fully implanted recording and stimulation systems that establish a direct link between cortical signals and the analogue modulation of epidural electrical stimulation targeting the spinal cord regions involved in the production of walking. A highly reliable BSI is calibrated within a few minutes. This reliability has remained stable over one year... This digital bridge establishes a framework to restore natural control of movement after paralysis." → AI Decoding Brainwaves Nature. 24 May 2023



<section-header><section-header><text>

"Using only EEG data, without participant specific fMRI-informed source analysis, we were able to identify the music a participant was listening to with a mean rank accuracy of 59.2%" → AI Decoding Brainwaves Daly, Ian. 12 January 2023





"175 volunteers recorded with magneto-encephalography or electro-encephalography while they listened to short stories and isolated sentences. The results show that our model can identify, from 3 seconds of magneto-encephalography signals, the corresponding speech segment with up to 41% accuracy out of more than 1,000 distinct possibilities on average across participants, and with up to 80% in the best participants—a performance that allows the decoding of words and phrases absent from the training set ."  $\rightarrow$  Al Decoding Brainwaves Meta. 05 October 2023

## DECODING SPEECH PERCEPTION FROM NON-INVASIVE BRAIN RECORDINGS











#### TOWARD A REAL-TIME DECODING OF IMAGES FROM BRAIN ACTIVITY

"Using magnetoencephalography (MEG), a non-invasive neuroimaging technique in which thousands of brain activity measurements are taken per second, we showcase an AI system capable of decoding the unfolding of visual representations in the brain with an unprecedented temporal resolution. ... Decoding of visual stimuli need not be restricted to a limited set of classes, but can now leverage pretrained representations to condition subsequent generative AI models. While the resulting image may be partly "hallucinated", interpreting images can be much simpler than interpreting latent features." → AI Decoding Brainwaves Meta. 18 October 2023



Clarinargsvisth, \* TIKIK



#### MOVING MAGNETOENCEPHAL OGRAPHY TOWARDS REAL-WORLD APPLICATIONS WITH A WEARABLE SYSTEM

"a magnetoencephalography system that can be worn like a helmet, allowing free and natural movement during scanning. This is possible owing to the integration of quantum sensors which do not rely on superconducting technology, with a system for nulling background magnetic fields. We demonstrate human electrophysiological measurement at millisecond resolution while subjects make natural movements, including head nodding, stretching, drinking and playing a ball game." → AI Decoding Brainwaves Nature. 21 March 2018





## CINEMATIC MINDSCAPES: HIGH-QUALITY VIDEO RECONSTRUCTION FROM BRAIN ACTIVITY

"We propose Mind-Video, which progressively learns spatiotemporal information from continuous fMRI data through masked brain modeling + multimodal contrastive learning + spatiotemporal attention + co-training with an augmented Stable Diffusion model that incorporates network temporal inflation." → AI Decoding Brainwaves NeurIPS. 19 May 2023



#### BUTTERFLY NETWORK ANNOUNCES 5-YEAR CO-DEVELOPMENT AGREEMENT WITH FOREST NEUROTECH FOR NEXT-GENERATION BRAIN COMPUTER INTERFACES USING ULTRASOUND-ON-CHIP<sup>TM</sup> TECHNOLOGY.

"Companies announce collaboration to develop the first implanted, whole-brain neural interface Powered by Butterfly's Ultrasound-on-Chip technology™" → Al Decoding Brainwaves Nature. 23 October 2023



## DECODING MOTOR PLANS USING A CLOSED LOOP ULTRASONIC BRAIN-MACHINE INTERFACE

"Brain-machine interfaces (BMIs) enable people living with chronic paralysis to control computers, robots and more with nothing but thought. Existing BMIs have trade-offs across invasiveness, performance, spatial coverage and spatiotemporal resolution. Functional ultrasound (fUS) neuroimaging is an emerging technology ... a new class of **less-invasive** (epidural) interfaces that generalize across extended time periods and promise to restore function to people with neurological impairments ... We streamed fUS data from the posterior parietal cortex of two rhesus macaque monkeys while they performed eye and hand movements. After training, the monkeys controlled up to eight movement directions using the BMI "  $\rightarrow$  AI Decoding Brainwaves

Nature. 30 November 2023 (received Jan 2023)



## NEURALINK

"The N1 Implant records neural activity through 1024 electrodes distributed across 64 threads."

"By modeling the relationship between different patterns of neural activity and intended movement directions, we can build a model (i.e., "calibrate a decoder") that can predict the direction and speed of an upcoming or intended movement."

→ AI Decoding Brainwaves

<u>Musk, Elon, and Neuralink. "An Integrated Brain-Machine Interface Platform with Thousands of</u> <u>Channels.</u>" bioRxiv, August 2, 2019. + <u>"Pager Plays MindPong" Neuralink Blog</u>. 2021

The <u>first human received an implant from @Neuralink</u> on Jan 28, 2024 March 20, 2024 Livestream of @Neuralink demonstrating "Telepathy " – controlling a computer and playing video games just by thinking

1.Quadriplegia (paralysis or severe weakness in all four
2.Paraplegia (paralysis in at least two limbs)
3.Vision loss
4.Hearing loss
5.Aphasia (inability to speak)
6.Amputee (major limb amputation)

limbs)



# 

FROM MESSY, MOIST, NOISY, NON-LINEAR, TURBULENT SYNAPTIC SIGNALS, AI-ENHANCED SENSORS GATHER, FILTER AND THEN PASS THE RELEVANT ASPECTS OF THE SIGNAL TO ANOTHER AI (WHICH EITHER ACTUATES A MOTOR NEURON, A DIFFUSION IMAGE GENERATOR, OR A LANGUAGE MODEL).

AFTER THE FILTRATION OF THE SIGNAL ARISING FROM THE BRAIN'S SIMULATION, CLEANED DATA IS PASSED TO A GENERATIVE MODEL, AND AN EXTERNAL REPRESENTATION OF WHAT WAS PREVIOUSLY INTERNAL ARISES. THIS IS AN ARTFUL ACT, THE MAKING OF A REPRESENTATION BASED UPON NEUROLOGICAL SIGNALS IN A DIRECT FEEDBACK PROCESS:

GATHER, FILTER, ACTUATE, REPEAT



# Al Designing Medicines

Hallucinated proteins, genetic errors, gene editors, vaccines, xrays, diagnostics, psychedelics

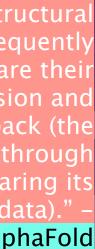


## ALPHAFOLD REVEALSTHE **STRUCTURE OF** THE PROTEIN UNIVERSE

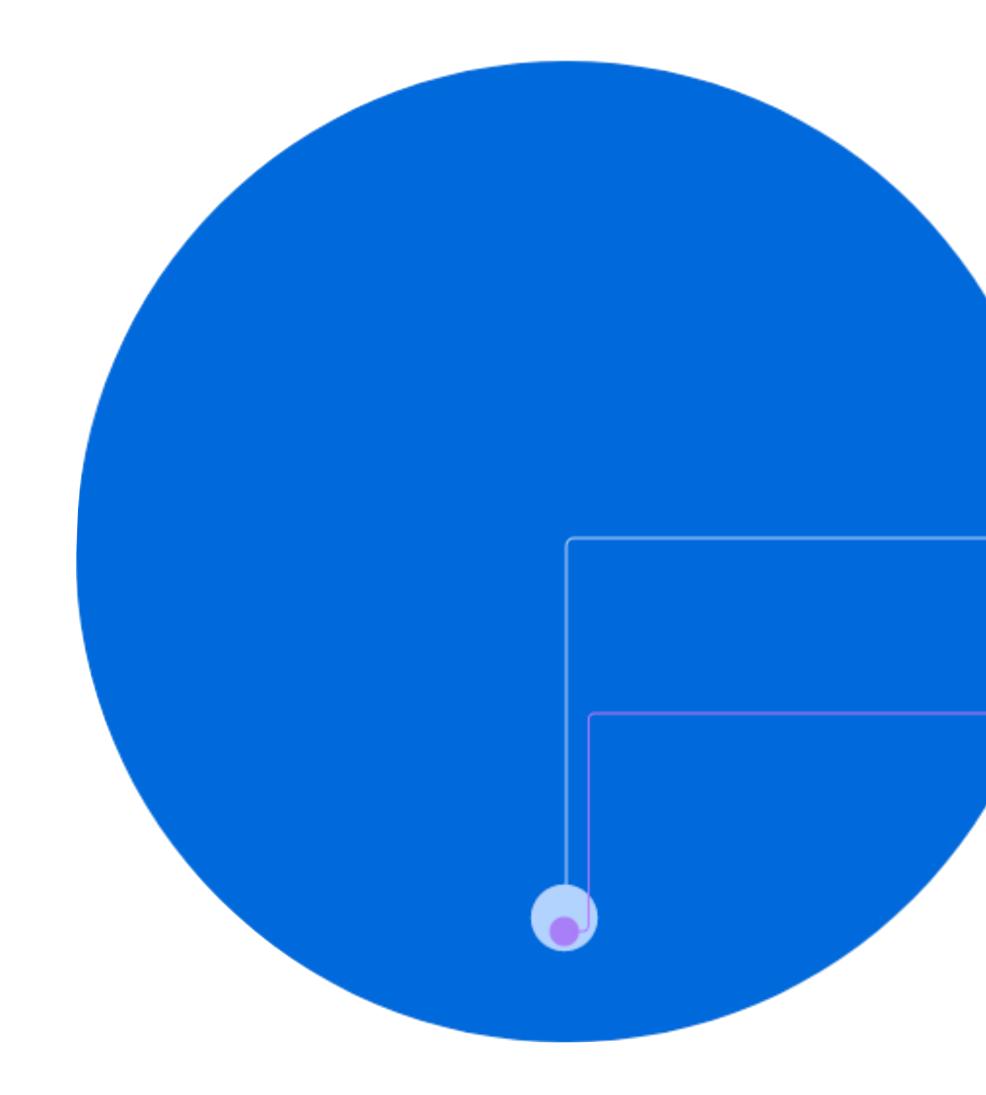
Released: "22 July, 2021 – covered over 350,000 structures, including the human proteome. 28 July, 2022, database expanded from nearly one million structures to over 200 million structures. AlphaFold Protein Structure Database (AlphaFold DB) to freely share this scientific knowledge with the world" → AI Designing Medicines DeepMind. 28 July, 2022

#### CARVING METAPHOR

carving (predicting specific structura elements of the protein), frequently tepping back to compare their progress with their initial vision and adjusting based on feedback (the model refines its predictions through terative processes, comparing its predictions with the training data). GPT4 describes AlphaFold



#### Number of Protein Structures



AlphaFold reveals the structure of the protein universe

Published **28 July 2022** Authors : **Demis Hassabis** 

AlphaFold DB today

200M+ Structures

AlphaFold DB previously

~1M Structures

Experimental (PDB) today 190K Structures

Epistemological explosion.

- Enzymatic catalysis,
- Transport and storage of molecules (e.g., oxygen transport by hemoglobin),
- Coordination of bodily functions (e.g., hormones like insulin),
- Mechanical support (e.g., actin and myosin in muscle fibers),
- Immune protection (e.g., antibodies),
- Generation and transmission of nerve impulses (e.g., neurotransmitter receptors),
- Control of growth and differentiation (e.g., growth factors),
- Movement (e.g., proteins in muscle contraction),
- Structural components (e.g., collagen in connective tissues),
- Signal transduction (e.g., receptor proteins),
- Cellular adhesion (e.g., integrins),
- Energy production and storage
- Histones. Around which genes are wrapped in the nucleosome: silencing or activating transcription of proteins from DNA

#### **1.Enzymatic catalysis in metabolic pathways:**

- pathway and the citric acid cycle (Krebs cycle) catalyze reactions that produce ATP, the primary energy currency of the cell.
- Enzymes like ATP synthase play a direct role in the production of ATP during oxidative phosphorylation in mitochondria.

#### **2.Transport and storage of molecules contributing to energy metabolism:**

- Hemoglobin transports oxygen from the lungs to tissues where it's used for oxidative phosphorylation, a process that generates ATP.
- Myoglobin in muscle cells stores oxygen and releases it during intense physical activity to meet increased energy demands.

#### **3.Regulation of energy homeostasis:**

- production.

#### **4.Structural roles related to energy storage and usage:**

- are directly involved in the mechanisms by which ATP is used to generate mechanical work.

• Proteins, particularly enzymes, are essential in metabolic pathways that convert nutrients into energy. For example, enzymes involved in the glycolysis

• Insulin, a hormone, plays a key role in regulating glucose levels in the blood and facilitates the uptake of glucose by cells, where it can be used for energy

• Other proteins act as receptors and signal transducers in pathways that control energy balance and metabolism, responding to nutritional states and energy needs.

• Actin and myosin are involved in muscle contraction, a process that consumes ATP. During physical activity, the demand for ATP increases, and these proteins

• Additionally, proteins can be a source of energy themselves; when other sources are depleted, proteins can be broken down into amino acids and converted into



### ALPHA MISSENSE (A VARIANT OF ALHPA FOLD)

"... categorised 89% of all 71 million possible genetic missense variants as either likely pathogenic or likely benign. By contrast, only 0.1% have been confirmed by human experts. All predictions made <u>freely available</u> to the research community and model code open sourced."  $\rightarrow$  Al Designing Medicines **DeepMind**. 19 September 2023



Tiny transcription errors in the amino acid sequences of fundamental proteins can lead to a wide range of dysfunctions and diseases, depending on the protein involved and the nature of the error. These errors can occur during DNA transcription (copying DNA into mRNA) or translation (converting mRNA into protein), leading to changes in the amino acid sequence of the resulting protein. The severity and type of dysfunction depend on how the error affects the protein's structure and function. Here are some examples of potential consequences:

- lead to the production of a malfunctioning protein involved in chloride ion transport.
- aggregates in the brain.
- **4.Metabolic Disorders**: Errors in enzymes can disrupt metabolic pathways, leading to the accumulation of toxic substances or deficiency of critical products. Phenylketonuria (PKU) is caused by mutations in the gene for the enzyme phenylalanine hydroxylase, leading to the accumulation of phenylalanine and subsequent brain damage if untreated.
- autoimmune diseases where the body attacks its own tissues.
- in skeletal, eye, and cardiovascular abnormalities.

These examples underscore the critical importance of accurate protein synthesis for health and the potential consequences of even minor errors in this process.

**1.Loss of Function**: If the error disrupts a critical area of the protein, it may no longer be able to perform its biological function, leading to diseases related to the loss of that function. For example, cystic fibrosis results from mutations in the CFTR gene that

**2.Gain of Function**: Errors can also result in a protein gaining a new, often harmful function. For example, certain mutations in the Ras gene lead to an active form of the Ras protein that continuously signals cell division, contributing to cancer development.

3.Misfolding and Aggregation: Incorrect amino acids can cause a protein to fold improperly, potentially leading to aggregation of misfolded proteins. Diseases such as Alzheimer's and Parkinson's are associated with the accumulation of misfolded protein

**5.Immune Response**: Some errors might create proteins that are recognized as foreign by the body's immune system, leading to

**6.Developmental Disorders**: Errors in proteins critical for development can lead to congenital abnormalities and developmental disorders. For example, mutations in the fibrillin-1 gene can lead to Marfan syndrome, affecting connective tissue and resulting



## SEARCH ALGORITHM REVEALS NEARLY 200 NEW KINDS OF CRISPR SYSTEMS

"To mine databases of protein and nucleic acid sequences for novel CRISPR systems, the researchers developed an algorithm based on an approach borrowed from the big data community. This technique, called locality-sensitive hashing, clusters together objects that are similar but not exactly identical." → Al Designing Medicines Broad Institute of MIT and Harvard. 23 November 2023 (Submitted almost a year earlier)

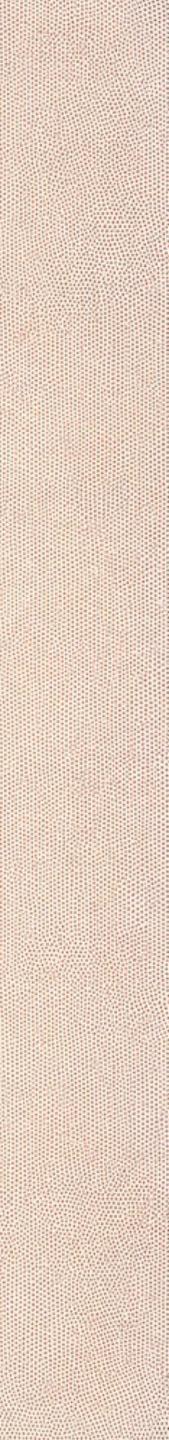
## DISCOVERY OF A STRUCTURAL CLASS OF ANTIBIOTICS WITH EXPLAINABLE DEEP LEARNING

"applied ensembles of graph neural networks to predict antibiotic activity and cytotoxicity for 12,076,365 compounds. ...empirically tested 283 compounds and found... one is selective against methicillin-resistant S. aureus (MRSA) ... and reduces bacterial titres in mouse models of MRSA skin and systemic thigh infection. Our approach enables the deep learning-guided discovery of structural classes of antibiotics and demonstrates that machine learning models in drug discovery can be explainable"  $\rightarrow$  AI Designing Medicines Nature. 20 December 2023



## ALGORITHM OPTIMIZED RNA DESIGN **IMPROVES** STABLITY AND IMMUNOGENIC

"LinearDesign substantially improves mRNA half-life and protein expression, and profoundly increases antibody titre by up to 128 times in mice compared to the codon-optimization benchmark on mRNA vaccines for COVID-19 and varicella-zoster virus"
 → Al Designing Medicines
 'Remarkable' Al tool designs mRNA vaccines that are more potent and stable
 Baidu. Nature. 02 May, 2023.



### LARGE-SCALE PANCREATIC CANCER **DETECTION VIA NON-CONTRASTCT AND DEEP LEARNING**

"... pancreatic cancer detection with artificial intelligence (PANDA), that can detect and classify pancreatic lesions with high accuracy via non-contrast CT. PANDA is trained on a dataset of 3,208 patients from a single center. PANDA ... outperforms the mean radiologist performance by 34.1% in sensitivity and 6.3% in specificity for PDAC identification, and achieves a sensitivity of 92.9% and specificity of 99.9% for lesion detection in a real-world multi-scenario validation consisting of 20,530 consecutive patients." → AI Designing Medicines EGNO MOD MOD Nature Medicine. 20 November 2023







### **ACOUSTIC ANALYSIS AND PREDICTION OF TYPE2 DIABETES MELLITUS USING SMARTPHONE-RECORDED VOICE SEGMENTS**

'267 participants diagnosed as nondiabetic (79 women and 113 men) or T2DM (18 women and 57 men) on the basis of American Diabetes Association guidelines were recruited in India between August 30, 2021 and June 30, 2022. Using a smartphone application, participants recorded a fixed phrase up to 6 times daily for 2 weeks, resulting in 18,465 recordings. ... Significant differences were found between voice recordings of nondiabetic and T2DM men and women... voice has emerged as a promising candidate for pathology detection and screening. It is noninvasive, inexpensive, and convenient." → AI Designing Medicines ScienceDirect. December 2023





### ALPHAFOLD FOUND THOUSANDS OF POSSIBLE PSYCHEDELICS. WILL ITS PREDICTIONS HELP DRUG DISCOVERY?

"Researchers have doubted how useful the AI protein-structure tool will be in discovering medicines — now they are learning how to deploy it effectively... AlphaFold structures identified the drugs that activated the **serotonin [ & G-protein-coupled] receptors** most potently. The psychedelic drug LSD works partly through this route, and many researchers are looking for non-hallucinogenic compounds that do the same thing, as potential antidepressants. ... GPCR are signaling pathways related to many diseases i.e. mental, metabolic including endocrinological disorders, immunological including viral infections, cardiovascular, inflammatory, senses disorders, and cancer." → AI Designing Medicines Nature. 20 January 2024



# -> AI MEDICINE CONCLUSIONS

MAPPING **PROTEIN** FOLDING BY IMAGINING BONDS AND TOPOLOGICAL PHYSICS. CATEGORIZING **GENETIC** *MISSENSE* **MUTATIONS** BY IMAGINING DISEASE OUTCOMES. DISCOVERING FOR **CRISPR SYSTEMS** BY RECOGNIZING SIMILARITIES WITHIN DATA HASHES. FINDING **NEW ANTIBIOTICS** BY PREDICTING ANTIBIOTIC ACTIVITY AND CYTOTOXICITY. OPTIMIZING **VACCINES** BY UNTANGLING PROTEINS. READING **X-RAYS**. DIAGNOSING **DIABETES** BY LISTENING TO THE VOICE. IDENTIFYING **DRUGS** THAT ACTIVATE SEROTONIN & G-PROTEIN-COUPLED] RECEPTORS.

IMPLICATIONS: AI SIMULATIONS OF DISEASE PROGRESSION THAT ARE CONFIGURED ON ACCURATE PERSONALIZED PROTEOMIC DATA WILL RECONFIGURE MEDICINE.

# 

#### Hallucinating de novo proteins, crystals, oral bacteria diets.

#### **FORCEGEN: END-TO-END DE NOVO PROTEIN GENERATION BASED ON NONLINEAR MECHANICAL UNFOLDING RESPONSES USING A LANGUAGE DIFFUSION MODEL**

"A generative language diffusion model can design proteins with desired nonlinear mechanical properties. Through evolution, nature has presented a set of remarkable protein materials, including elastins, silks, keratins and collagens with superior mechanical performances that play crucial roles in mechanobiology. Via full-atom molecular simulations for direct validation, we demonstrate that the designed proteins are de novo, and fulfill the targeted mechanical properties, including unfolding energy and mechanical strength, as well as the detailed unfolding force-separation curves. Our model offers rapid pathways to explore the enormous mechanobiological protein sequence space unconstrained by biological synthesis, using mechanical features as the target to enable the discovery of protein materials with superior mechanical properties."

→ AI Material Discovery Science Advances. 7 February, 2024

## GRAPH NETWORKS FOR MATERIALS EXPLORATION (GNOME)

"finds 2.2 million new crystals, including 380,000 stable materials – equivalent to nearly 800 years' worth of knowledge. **GNOME**'s predictions are available to the research community: contributing 380,000 materials predicted to be stable to the Materials Project" → Al Material Science DeepMind. 29 November 2023



Semiconductors, integrated circuits, microchips Piezoelectronics cellphones, watches, anything that has a clock cycle Sensors, actuators, ultrasound imaging devices **LCD** Liquid crystal displays Lasers: surgical, etc Solar panels (crystalline silicon photovoltaic cells) Crystal oscillators: data transmission, GPS, aerospace Laser doubling: Medical diagnostic, quantum physics Thermal conductivity application: heat sinks and cutting tools Intercalcation electrodes: rechargeable batteries

# CRYSTALS

## AN AUTONOMOUS LABORATORY FOR THE ACCELERATED SYNTHESIS OF NOVEL MATERIALS

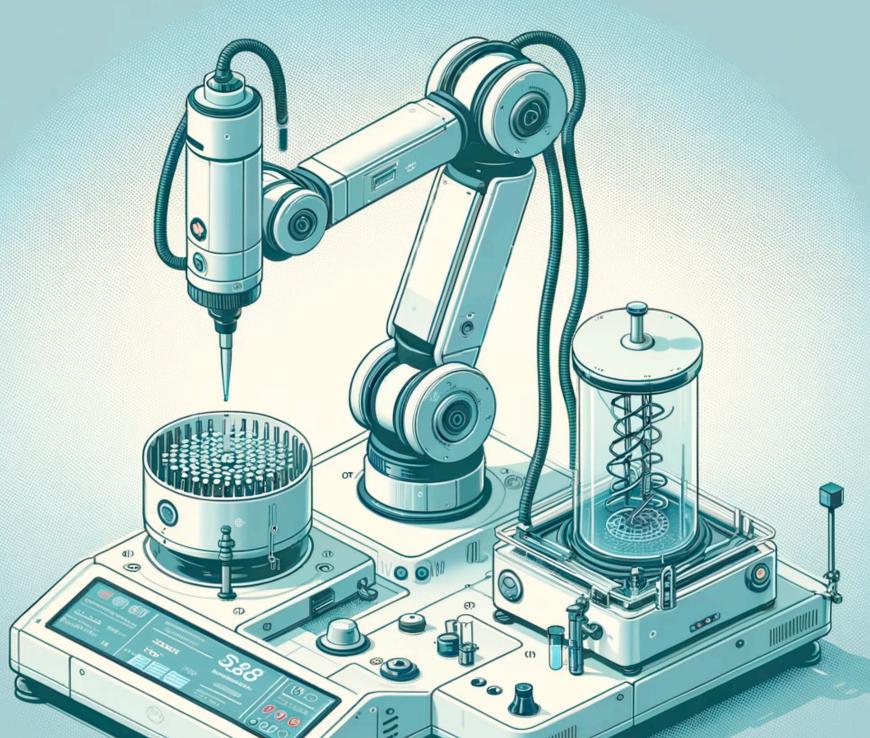
"Over 17 days of continuous operation, the A-Lab realized 41 novel compounds from a set of 58 targets including a variety of oxides and phosphates that were identified using large-scale ab initio phase-stability data from the Materials Project and Google DeepMind. Synthesis recipes were proposed by natural-language models trained on the literature and optimized using an active-learning approach grounded in thermodynamics. Analysis of the failed syntheses provides direct and actionable suggestions to improve current techniques for materials screening and synthesis design. The high success rate demonstrates the effectiveness of artificial-intelligence-driven platforms for autonomous materials discovery and motivates further integration of computations, historical knowledge and robotics." → Al Material Science

Lawrence Berkeley National Laboratory 29 November 2023



## AUTONOMOUS CHEMICAL RESEARCH WITHLLMS

"Coscientist, an artificial intelligence system driven by GPT-4 that autonomously designs, plans and performs complex experiments by incorporating large language models empowered by tools such as internet and documentation search, code execution and experimental automation. Coscientist showcases its potential for accelerating research across six diverse tasks, including the successful reaction optimization of palladium-catalysed cross-couplings, while exhibiting advanced capabilities for (semi-)autonomous experimental design and execution." → Al Material Science Nature. December 20, 2023





#### BACTERAI MAPS MICROBIAL MICROBIAL METABOLISM WITHOUT PRIOR NOWLEDGE

"an automated science platform that maps microbial metabolism but requires no prior knowledge. BacterAl learns by converting scientific questions into simple games that it plays with laboratory robots. The agent then distils its findings into logical rules that can be interpreted by human scientists. We use BacterAl to learn the amino acid requirements for two oral streptococci: Streptococcus gordonii and Streptococcus sanguinis. We then show how transfer learning can accelerate BacterAl when investigating new environments or larger media with up to 39 ingredients. Scientific gameplay and BacterAl enable the unbiased, autonomous study of organisms for which no training data exist." → Al Robots

University of Michigan. **Nature**. 4 May 2023







# -> AIMATERIALS CONCLUSIONS

**PROMPTING PROTEINS** VIA FULL-ATOM MOLECULAR SIMULATION TO EXPLORE THE ENORMOUS MECHANOBIOLOGICAL PROTEIN SEQUENCE SPACE UNCONSTRAINED BY BIOLOGICAL SYNTHESIS.

#### FINDING 2.2 MILLION NEW CRYSTALS, THEN FILTERING TO 380,000 STABLE MATERIALS.

SYNTHESIZING NOVEL COMPOUNDS WITH SEMI-AUTONOMOUS ROBOTS USING RECIPES PROPOSED BY NATURAL-LANGUAGE MODELS, OPTIMIZED WITH **ACTIVE-LEARNING.** 

**IMPLICATIONS: MATERIAL EN IMPROVEMENT CYCLES:** 

HALLUCINATE, FILTER, SYNTHESIZE, OPTIMIZE, REPEAT

## 

#### Hallucinating rewards, names, depth, tasks, races, and factories.





#### EUREKA: HUMAN-LEVEL REWARD DESIGN VIA CODING LLMS

"Eureka exploits the remarkable zero-shot generation, code-writing, and in-context improvement capabilities of state-of-the-art LLMs, such as GPT-4, to perform evolutionary optimization over reward code. The resulting rewards can then be used to acquire complex skills via reinforcement learning. Without any task-specific prompting or pre-defined reward templates, Eureka generates reward functions that outperform expert human-engineered rewards. In a diverse suite of 29 open-source RL environments that include 10 distinct robot morphologies, **Eureka outperforms human experts on 83% of the tasks**, leading to an average normalized improvement of 52%."

 $\rightarrow$  AI Robots

NVIDIA UPenn. 19 October 2023



Image segmenatiion

Ball-iainbow

SEGMENT ANYTHING

"a new task, model, and dataset for image segmentation.... often competitive with or even superior to prior fully supervised results. We are releasing the Segment Anything Model (SAM) and corresponding dataset (SA-1B) of 1B masks and 11M images at <u>https://segment-anything.com/</u> to foster research into foundation models for computer vision."

→ Al Robots
Meta. 5 April 2023



#### DEPTHANY THIS UNLEASHING THE POWER OF LARGE-SCALL UNLABELE DATE

"a highly practical solution for robust monocular depth estimation. Without pursuing novel technical modules, we aim to build a simple yet powerful foundation model dealing with any images under any circumstances. To this end, we scale up the dataset by designing a data engine to collect and automatically annotate large-scale unlabeled data (~62M) ... models released: https://github.com/LiheYoung/Depth-Anything

→ Al Robots
ByteDance. 19 January 2024



## OPEN X-EMBODIMENT: ROBOTIC LEARNING DATASETS AND RT-X MODELS

"a dataset from 22 different robots collected through a collaboration between 21 institutions, demonstrating 527 skills (160,266 tasks). We show that a high-capacity model trained on this data, which we call RT-X, exhibits positive transfer and improves the capabilities of multiple robots by leveraging experience from other platforms."  $\rightarrow$  Al Robots

DeepMind + 34 international robotics labs. 13 Oct, 2023



0

### CHAMPION-LEVEL DRONE RACING USING DEEP REINFORCEMENT LEARNING

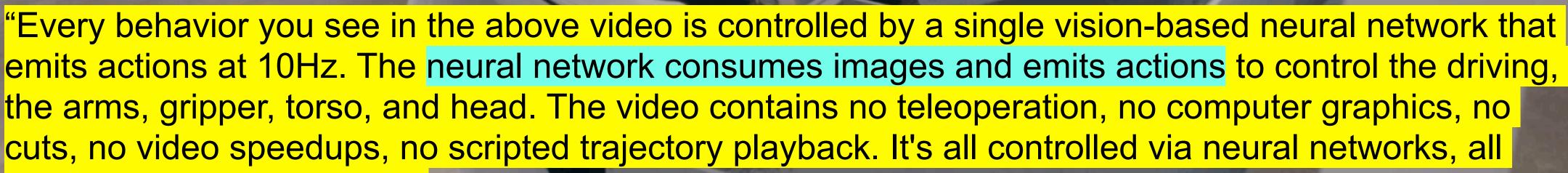
"Swift, an autonomous system that can race physical vehicles at the level of the human world champions. The system combines deep reinforcement learning (RL) in simulation with data collected in the physical world. Swift competed against three human champions, including the world champions of two international leagues, in real-world head-to-head races. Swift won several races against each of the human champions and demonstrated the fastest recorded race time"  $\rightarrow$  Al Robots Nature. 30 August, 2023.



## **1X EVE**

the arms, gripper, torso, and head. The video contains no teleoperation, no computer graphics, no cuts, no video speedups, no scripted trajectory playback. It's all controlled via neural networks, all autonomous, all 1X speed."

→ AI Robots **1X**. 8 February, 2024







## -> A ROBOTS CONCLUSIONS

#### EUREKA WRITES THE REWARDS THAT OPTIMIZES THE ROBOTS (WHO WILL CREATE THE ROBOTS).

THE **DEPTH-ANYTHING** AND THE **SEGMENT-ANYTHING** MODELS ARE BASICALLY ACCURATE, FAST FILTRATION SYSTEMS, BASED UPON A MASSIVE VISUAL EXPERIENTIAL DATABASE. THE DORSAL NEUROANATOMY OF ROBOTICS IS **BEING SOLVED.** 

NON-HUMANOID ROBOTS (SUCH AS DRONES) ARE ALREADY OPERATING AUTONOMOUSLY AT EXPERT LEVEL.

COLLABORATIVE, COLLECTIVE AGGREGATES OF ROBOTIC DATABASES FLOWING ACROSS DIFFERENT ROBOT MORPHOLOGIES SUGGEST THE EMERGENCE OF A GESTURE-ANYTHING, A DO-ANYTHING NEURAL NETWORK MODEL, WHICH IS FAST AND ACCURATE IN DECIDING APPROPRIATE ACTIONS IN CONTEXTUALLY NUANCED SITUATIONS.



# 

#### Hallucinated distinctions, discoveries, geometry.



## BRAIN ORGANOID RESERVOIR COMPUTING **FOR AI**

"adaptive reservoir computation of biological neural networks in a brain organoid. In this approach—which is termed Brainoware -computation is performed by sending and receiving information from the brain organoid using a high-density multielectrode array. By applying spatiotemporal electrical stimulation, nonlinear dynamics and fading memory properties are achieved, as well as unsupervised learning from training data by reshaping the organoid functional connectivity." → AI Brains

Nature Electronics. December 11, 2023





## FUNSEARCH: MAKING NEW DISCOVERIES IN MATHEMATICAL SCIENCES USING LLMS

"a method to search for new solutions in mathematics and computer science. FunSearch works by pairing a pre-trained LLM, whose goal is to provide creative solutions in the form of computer code, with an automated "evaluator", which guards against hallucinations and incorrect ideas... FunSearch discovered new solutions for the **cap set problem**, a longstanding open problem in mathematics. In addition, to demonstrate the practical usefulness of FunSearch, we used it to discover more effective algorithms for the "**bin-packing" problem**."  $\rightarrow$  Al Brains

Google DeepMind. December 14, 2023



"As LLMs have been shown to "hallucinate" factually incorrect information, using them to make verifiably correct discoveries is a challenge. But what if we could harness the creativity of LLMs by identifying and building upon only their very best ideas?"

#### ALPHA GEOMETRY

"AlphaGeometry is a neuro-symbolic system made up of a neural language model and a symbolic deduction engine, which work together to find proofs for complex geometry theorems. Akin to the idea of "thinking, fast and slow", one system provides fast, "intuitive" ideas, and the other, more deliberate, rational decision-making... solves complex geometry problems at a level approaching a human Olympiad gold-medalist ... and by developing a method to generate a vast pool of synthetic training data - 100 million unique examples - we can train AlphaGeometry without any human demonstrations, sidestepping the data bottleneck."

→ Al Brains
DeepMind. 17 January 2024



#### INDUCE + DEDUCE

9

0

0

1 Cr

# SINULATE THEN TEST

#### **NEURO+ SYMBOLIC**

0

#### SCALABLE INSTRUCTABLE MULTIWORLD AGENT (SIMA)

'SIMA, short for Scalable Instructable Multiworld Agent, a generalist AI agent for 3D virtual settings. We partnered with game developers to train SIMA on a variety of video games. This research marks the first time an agent has demonstrated it can understand a broad range of gaming worlds, and follow natural-language instructions to carry out tasks within them, as a human might. ... We hope that SIMA and other agent research can use video games as sandboxes to better understand how AI systems may become more helpful... The current version of SIMA is evaluated across 600 basic skills, spanning navigation (e.g. "turn left"), object interaction ("climb the ladder"), and menu use ("open the map")... an agent trained on many games was better than an agent that learned how to play just one. In our evaluations, SIMA agents trained on a set of nine 3D games from our portfolio significantly outperformed all specialized agents trained solely on each individual one. What's more, an agent trained in all but one game performed nearly as well on that unseen game as an agent trained specifically on it, on average. Importantly, this ability to function in brand new environments highlights SIMA's ability to generalize beyond its training." → Al Brains

DeepMind. 13 March 2024



# -> AI BRAINS CONCLUSIONS

BRAINOWARE: ORGANOIDS CONNECTED TO INTEGRATED CIRCUITS. FUN SEARCH: A FUNDAMENTAL SEARCH AGENT THAT HALLUCINATES THEN EVALUATES. ALPHAGEOMETRY: A NEURO-SYMBOLIC SYSTEM MADE UP OF A NEURAL LANGUAGE MODEL AND A SYMBOLIC DEDUCTION ENGINE, WHICH FINDS PROOFS FOR COMPLEX GEOMETRY THEOREMS. SIMA (SCALABLE INSTRUCTABLE MULTIWORLD AGENT): A GENERALIST AI AGENT FOR 3D VIRTUAL SETTINGS.

IMPLICATION: HUMANS HAVE OFTEN ASSUMED THEY ARE THE ONLY SENTIENT, CONSCIOUS, AGENTS ON THIS PLANET. RATIONAL CATEGORIES OF BEHAVIOR AND THINKING ARE NOW BEING SYNTHETICALLY DERIVED BY ALLOWING THE GENERATIVE AGILITY OF HALLUCINATORY AI PROCESSES TO FREELY AND CREATIVELY FLOW.

## CONCLUSIONS

#### CONTINGENT, SPECULATIVE, EPHEMERAL & BEAUTIFUL

-



## HALLUCINATIONS ARE NECESSARY & INNATE

TO AI, SCIENCE, REALITY, MINDS, SOCIETIES, PROTEINS, MAMMALS, & SYNTHETIC INTELLIGENT DISCOVERY

# SEMI-AUTONOMOUS EVOLVING MULTIMODAL INDUCTIVE-DEDUCTIVE AI RECURSIVE RESEARCH AGENTS ARE ARISING

#### rompting Let's Verify Step by Step

ML Placer: a Graph Placement Methodology for Fast Chip Design 🗨 Alpha Dev, Alpha Tensor, Alpha Code 2 Sakana Evolutionary Al

Decoding Speech Perception From Non-Invasive Brain Recordings Toward a Real-Time Decoding of Images From Brain Activity Cinematic Mindscapes: High-Quality Video Reconstruction From Brain Activity

**GNOME**: Graph Networks for Materials Exploration

Eureka: Human-Level Reward Design via Coding LlMs Segment Anything **Depth Anything Open X-Embodiment:** Robotic Learning Datasets and RT-X Models BacterAI: Maps Microbial Metabolism Without Prior Knowledge Champion-Level Drone Racing Using Deep Reinforcement Learning 1X Eve

Alpha Geometry SIMA

## RECAP

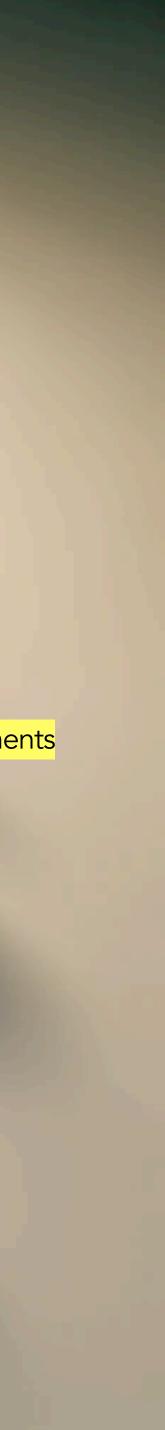
Organoids Geometry

AlphaFold Reveals the Structure of the Protein Universe Alpha Missense (a Variant of Alpha Fold)

Search Algorithm Reveals Nearly 200 New Kinds of CRISPR Systems Discovery of a Structural Class of Antibiotics With Explainable Deep Learning LinearDesign: Algorithm for Optimized mRNA Design Improves Stability and Immunogenicity PANDA Large-Scale Pancreatic Cancer Detection via Non-Contrast CT and Deep Learning Acoustic Analysis and Prediction of Type 2 Diabetes Mellitus Using Smartphone-Recorded Voice Segments AlphaFold Found Thousands of Possible **Psychedelics (Serotonin & G-Protein Coupled Recptors)** 

Brain Organoid Reservoir Computing for Al

FunSearch: Making New Discoveries in Mathematical Sciences Using LLMs



The preceding framework and examples become the foundation from which it is possible to claim: **If simulation underlies cognition**, and cognition tends toward boundary configurations that give rise to the seemingness of a self and the sense of consciousness, then current AI may internally emulate incipient self-identity formations.



#### Hallucinations are what we are. As is everything else.

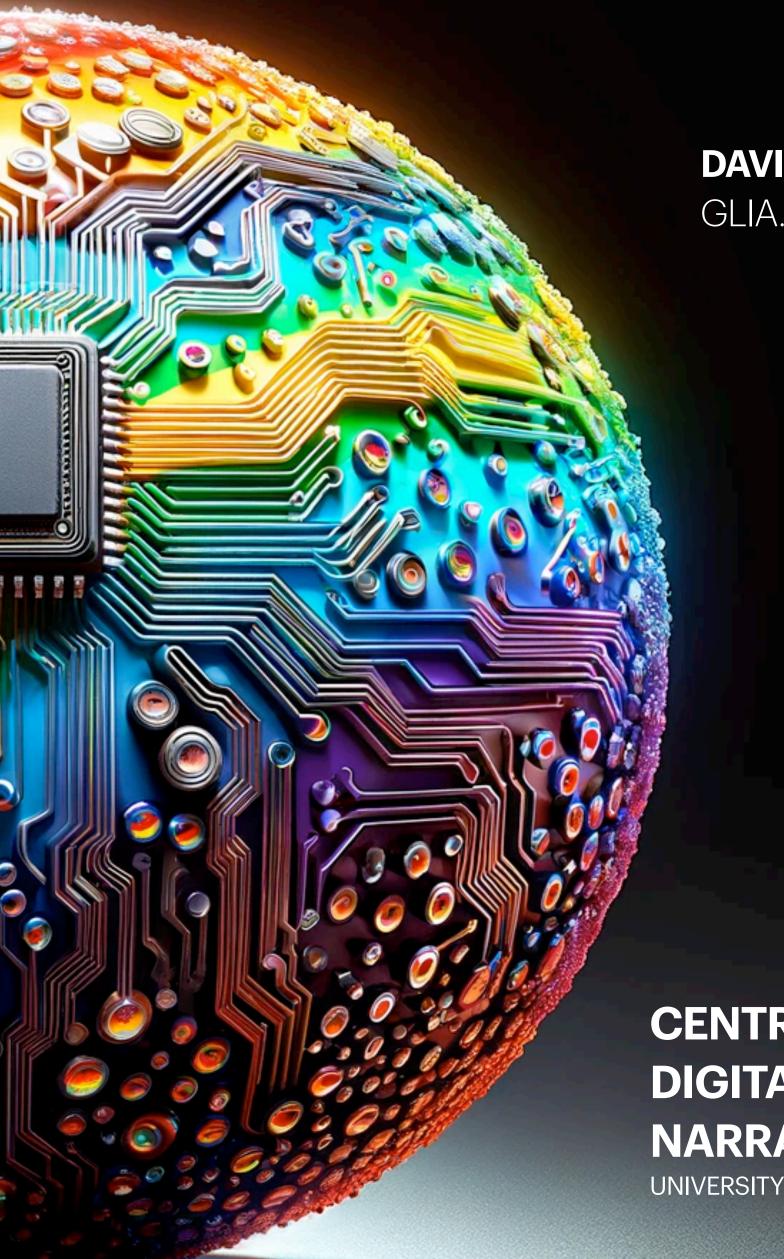
## CONCLUSION

Recursive hyper-accelerated semi-autonomous AI hallucinations will transfigure society.

1

## SPRING

AN ONGOING SET OF LINKS JHAVE.SUBSTACK.COM



**DAVID (JHAVE) JOHNSTON** GLIA.CA

CENTRE FOR DIGITAL NARRATIVE UNIVERSITY OF BERGEN



