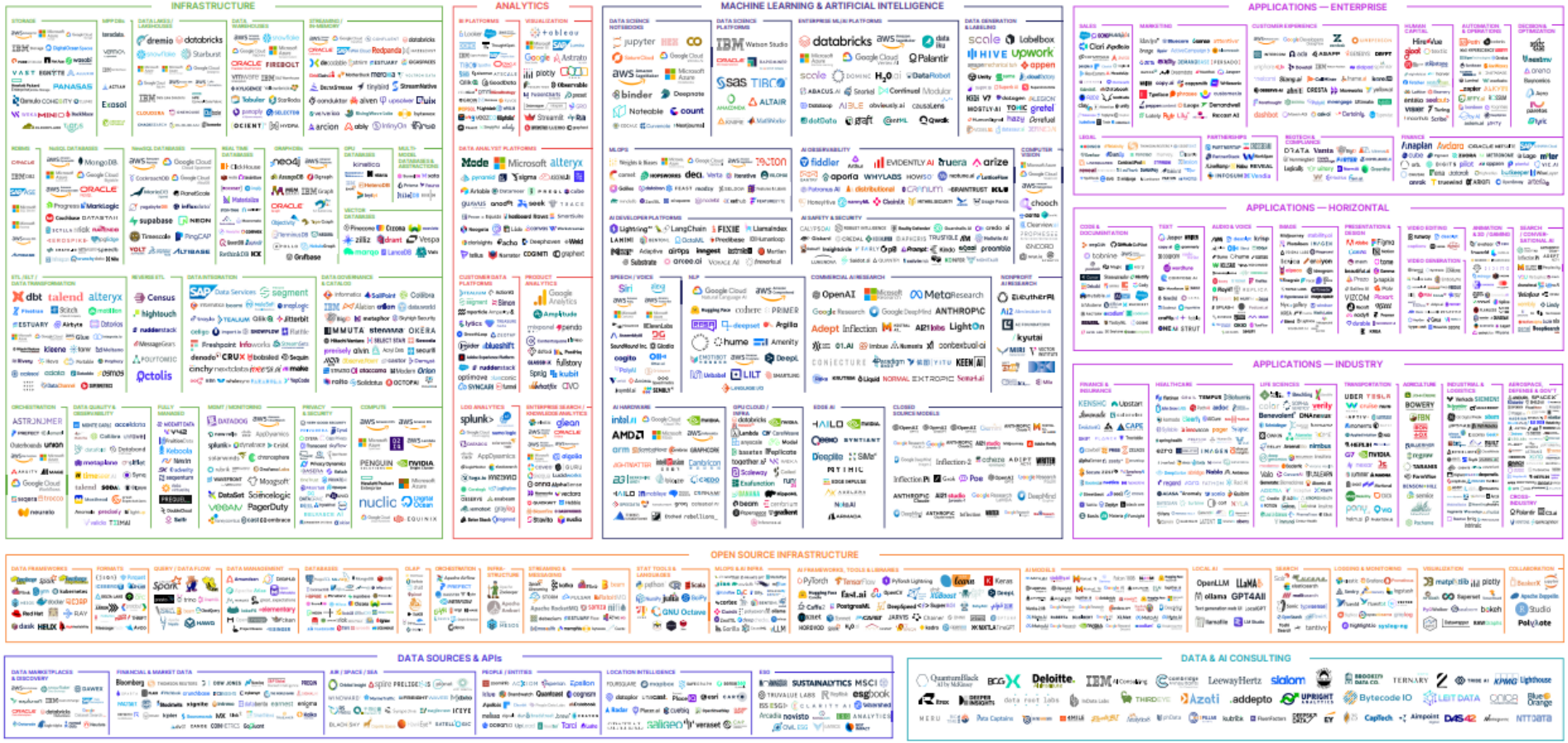
KI

Sabine Proßnegg und John Feiner

Version 1.0, Oktober 2025

THE 2024 MAD (MACHINE LEARNING, ARTIFICIAL INTELLIGENCE & DATA) LANDSCAPE

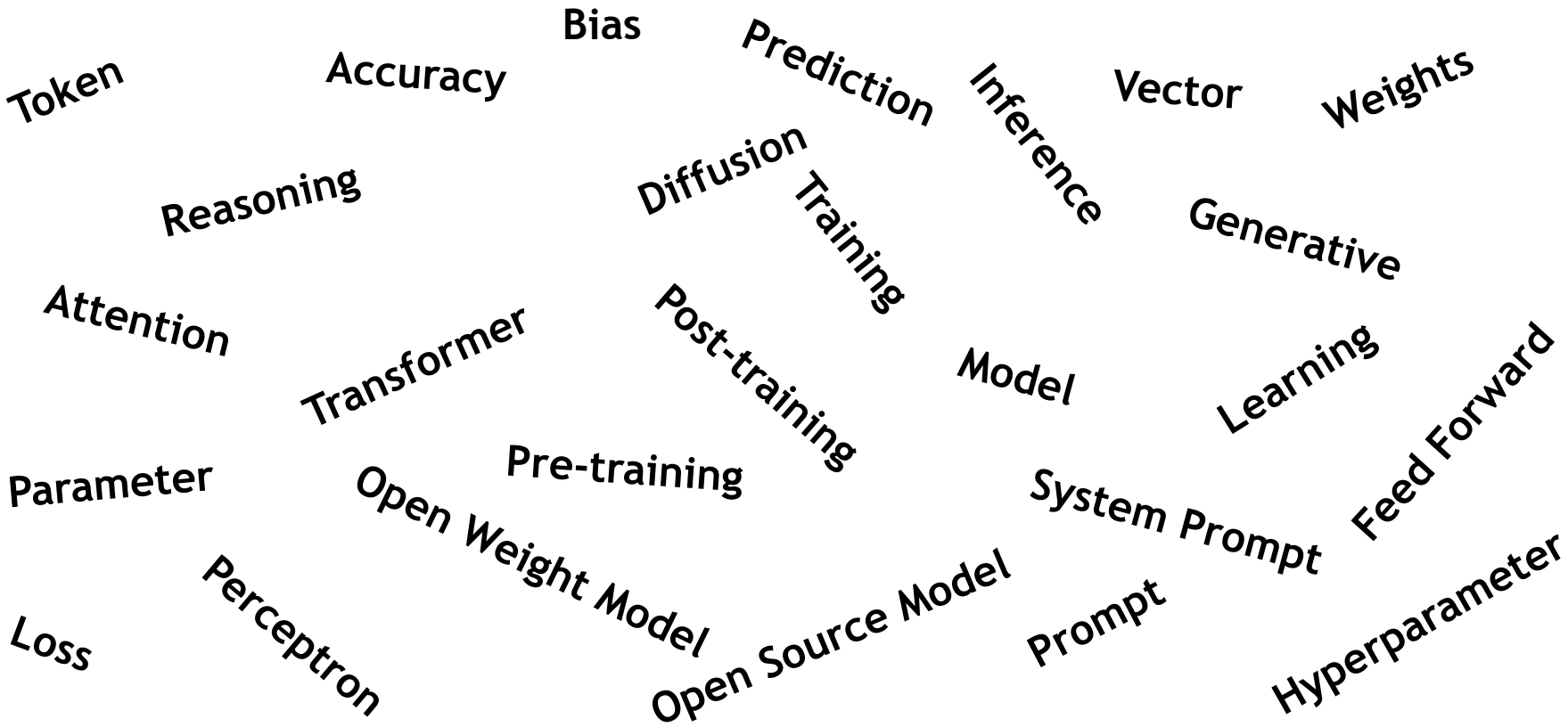


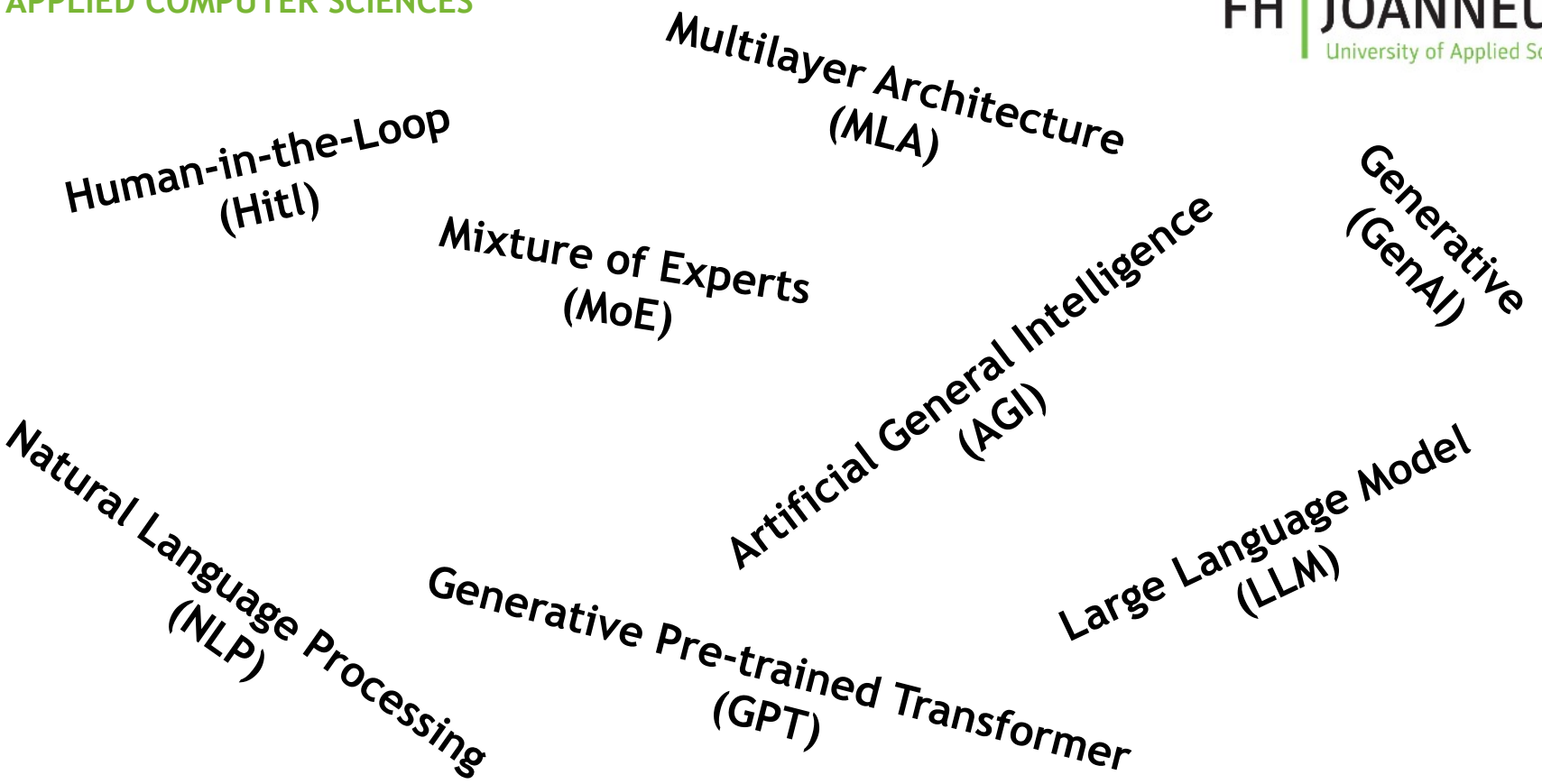
Version 1.0 – March 2024

Blog post: mattturck.com/MAD2024
Interactive version: MAD.firstmarkcap.com
Comments? Email MAD2024@firstmarkcap.com

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

https://mad.firstmark.com/


Bias
Token
Accuracy
Prediction
Inference
Vector
Weights
Reasoning
Diffusion
Training
Generative
Attention
Transformer
Post-training
Model
Learning
Feed Forward
Parameter
Pre-training
System Prompt
Open Weight Model
Perceptron
Open Source Model
Prompt
Hyperparameter
Loss


Multilayer Architecture
(MLA)
Human-in-the-Loop
(Hitl)
Mixture of Experts
(MoE)
Generative
(GenAI)
Artificial General Intelligence
(AGI)
Natural Language Processing
(NLP)
Large Language Model
(LLM)
Generative Pre-trained Transformer
(GPT)

Entstehungsgeschichte

Geschichte sehr(!) vereinfacht

1950 Turing: "**Turing Test**"
1965 Weizenbaum: **Elisa** (NLP) Chat
1997 Deep Blue: Gewinnt im **Schach**
2011 Watson: Gewinnt "**Jeopardy**!"
2017 AlphaGo Zero: Lernt sich selbst **Go**
2020 ChatGPT: Einfachst **benutzbarer** Chat

https://www.linkedin.com/pulse/inception-artificial-intelligence-ibm-watson-its-win-rajwanshi-

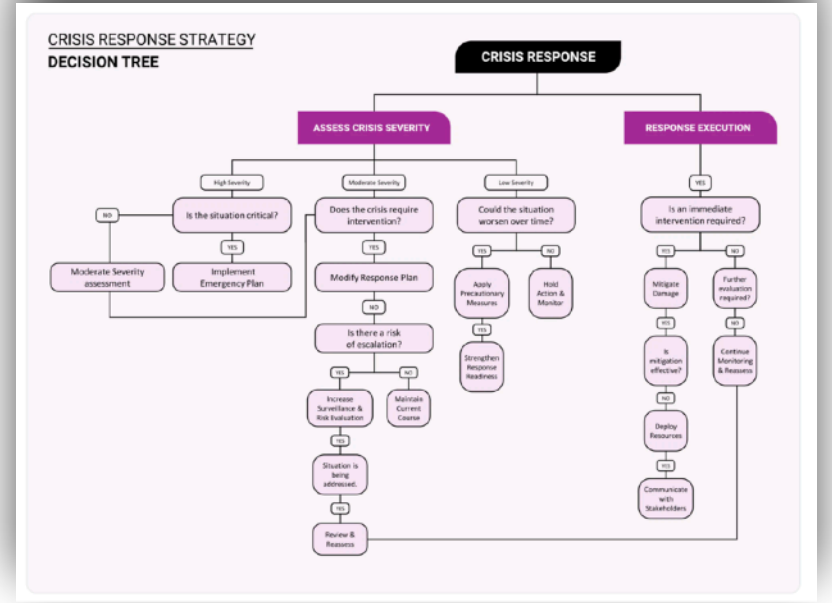
Wie funktioniert was?

Einige ausgewählte Begriffe / Beispiele

Früher üblich:

Regelbasierte Systeme

Programmierte Logik

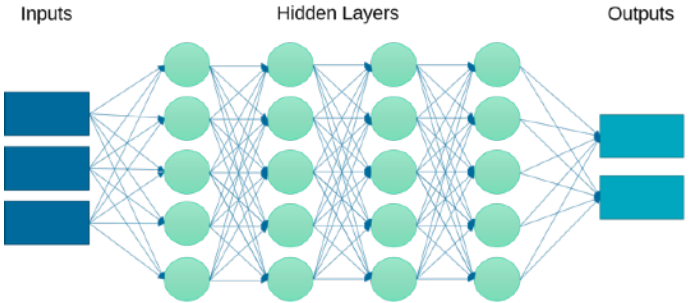


https://www.highfile.com/crisis-response-decision-tree-template/#google_vignette

Dann kamen:

Modelle mit

Trainierte "Logik"

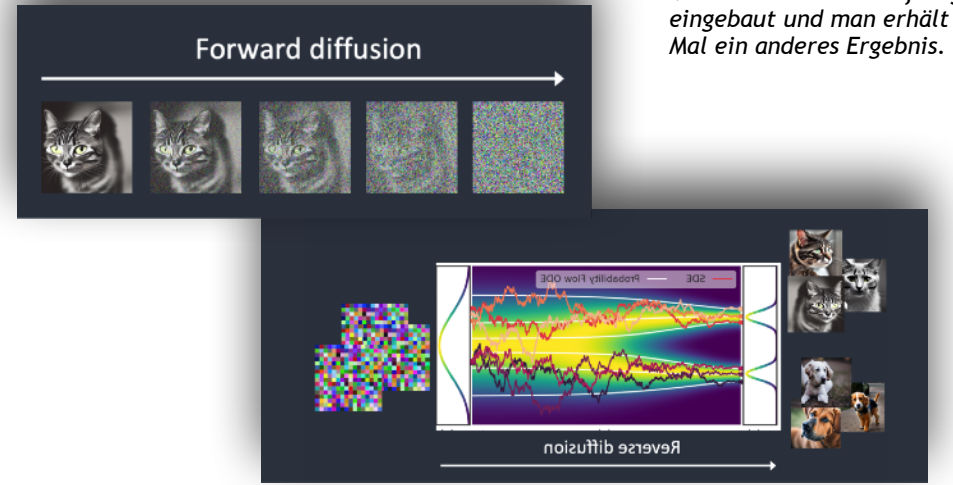


http://apmonitor.com/do/uploads/Main/deep_neural_network.png

Beispiel Bild-Klassifikation: Beim Trainieren (viele Bilder von Obst) wird eine innere Konfiguration optimiert, damit man einem bekanntem Input (Apfel) einem erwarteter Output (das ist ein Apfel) nahekommt. Dazu sind vielfache Durchläufe mit zuvor kategorisierten Daten (Viele Beispielbilder von je Äpfel, Bananen) nötig.

Schickt man nun ein neues Bild (Obstkorb, Schwedenbombe, ...) in das Modell so erhält man Wahrscheinlichkeiten als Resultat: das ist zu 69%, 13%, ... ein Apfel.

Diffusion - Bilder erfinden

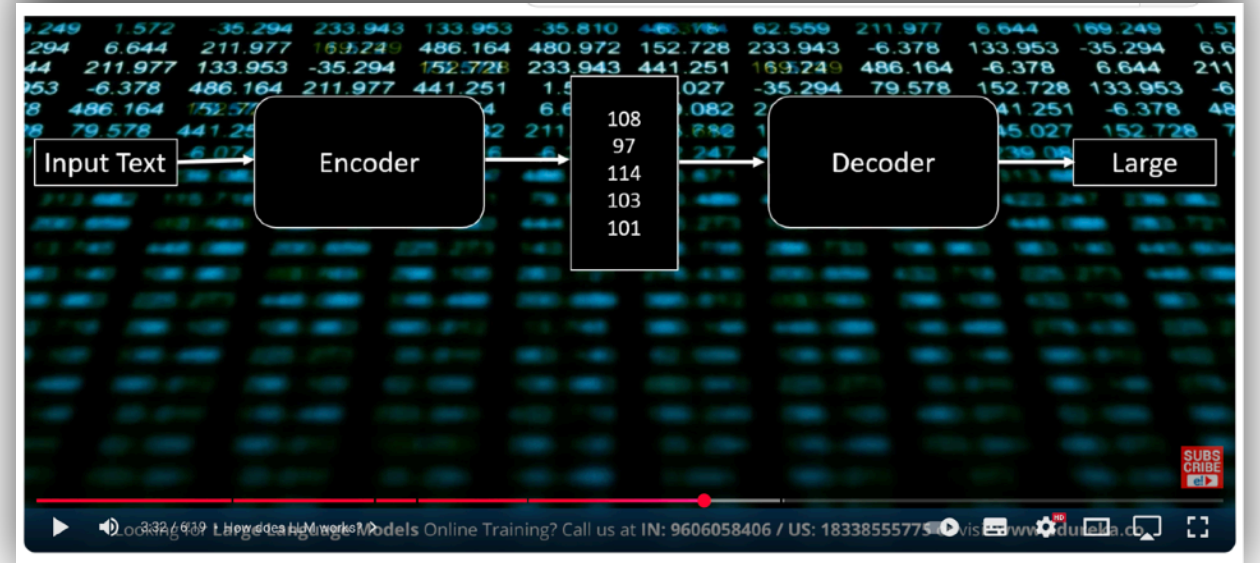


Mit Absicht wurde "Zufälligkeit" eingebaut und man erhält jedes Mal ein anderes Ergebnis.

https://stable-diffusion-art.com/how-stable-diffusion-work/

https://en.wikipedia.org/wiki/Stable_Diffusion

LLM ... Large Language Model



Auf Text spezialisiertes Modell.
Idee: mit aller Textinformation aus dem "gesamten Internet" trainiert.

https://www.youtube.com/watch?v=fn9T5yPyv1o

Bahnbrechendes Paper

+ Viele Daten

+ Rechenkapazität

+ ...

+ Optimierungen:
  "Attention is all you need"

https://www.datacenterdynamics.com/en/news/openai-and-oracle-to-deploy-450000-gb200-gpus-at-stargate-abilene-data-center/

Ausgewählte Probleme

Panta Rhei — alles fließt

Entscheidungen von AI Systemen

... sind nicht transparent und liefern

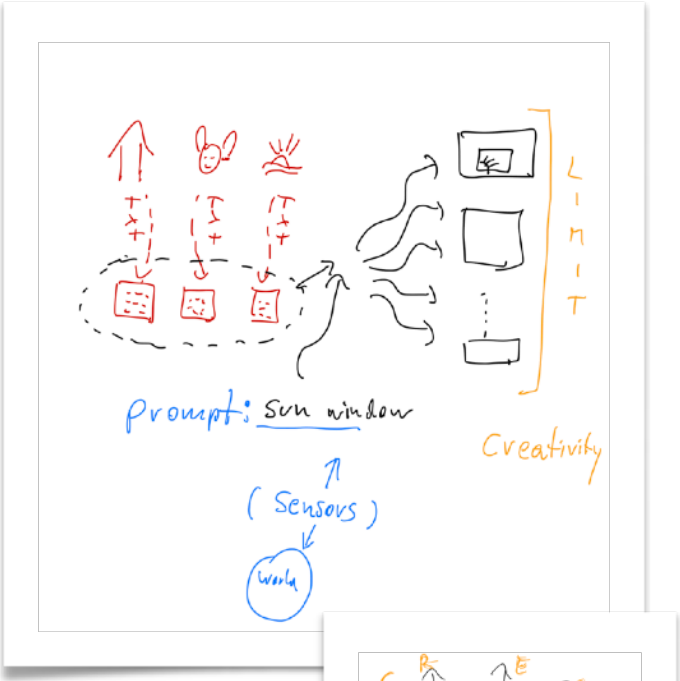
... jeweils **unterschiedliche** Ergebnisse

AI Systeme **ändern sich** ständig

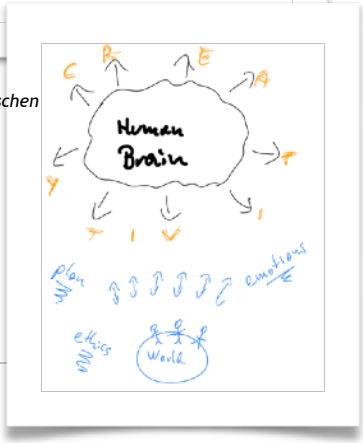
... Daten werden hinzugefügt

... Inklusive Benutzer:innen Eingaben (prompts)

... Modelle werde wiederholt trainiert



Vergleich Kreativität von Computersystemen und von Menschen



Vergessen

Nötig wäre:

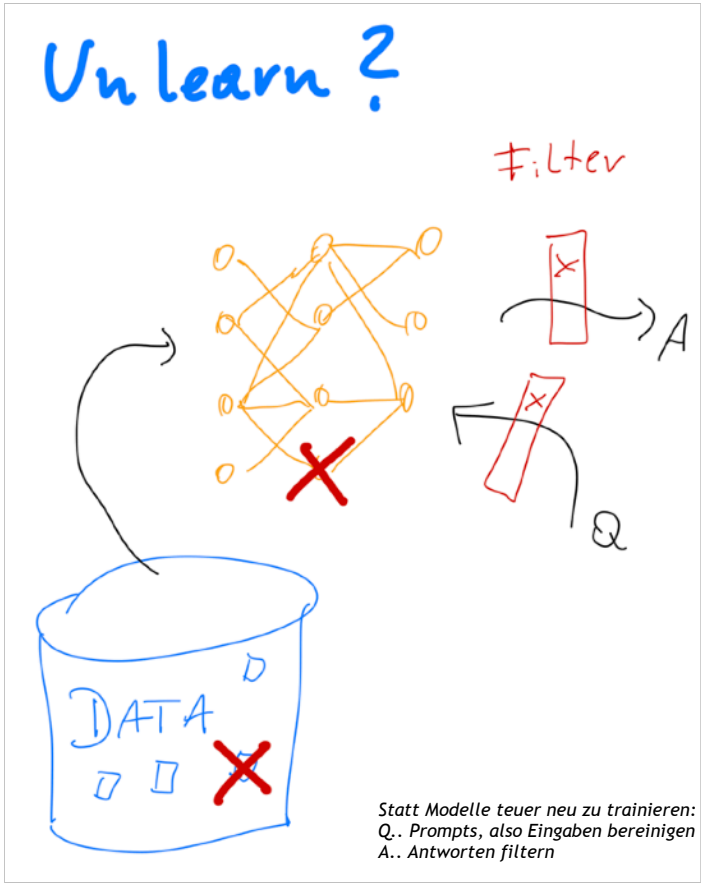
... Datensätze löschen

... Dann neu trainieren

Workaround:

... Eingaben filtern,

... Ausgaben nachbearbeiten



Statt Modelle teuer neu zu trainieren:
Q.. Prompts, also Eingaben bereinigen
A.. Antworten filtern

Probleme und Attacken

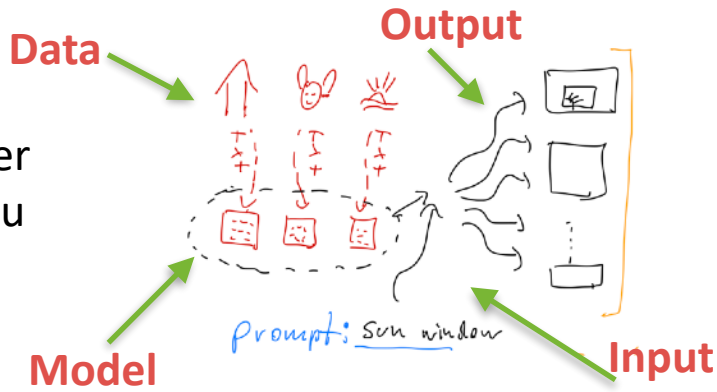
Besser "das Böse" erwarten:

... KI Systeme arbeiten unabsichtlich aber oft auch absichtlich falsch bzw. sind voreingenommen

Viele Wege Systeme zu "hacken":

... Ziel ist einfach nur Daten zu stehlen oder aber auch Eingaben, Modelle, Ausgaben zu verändern

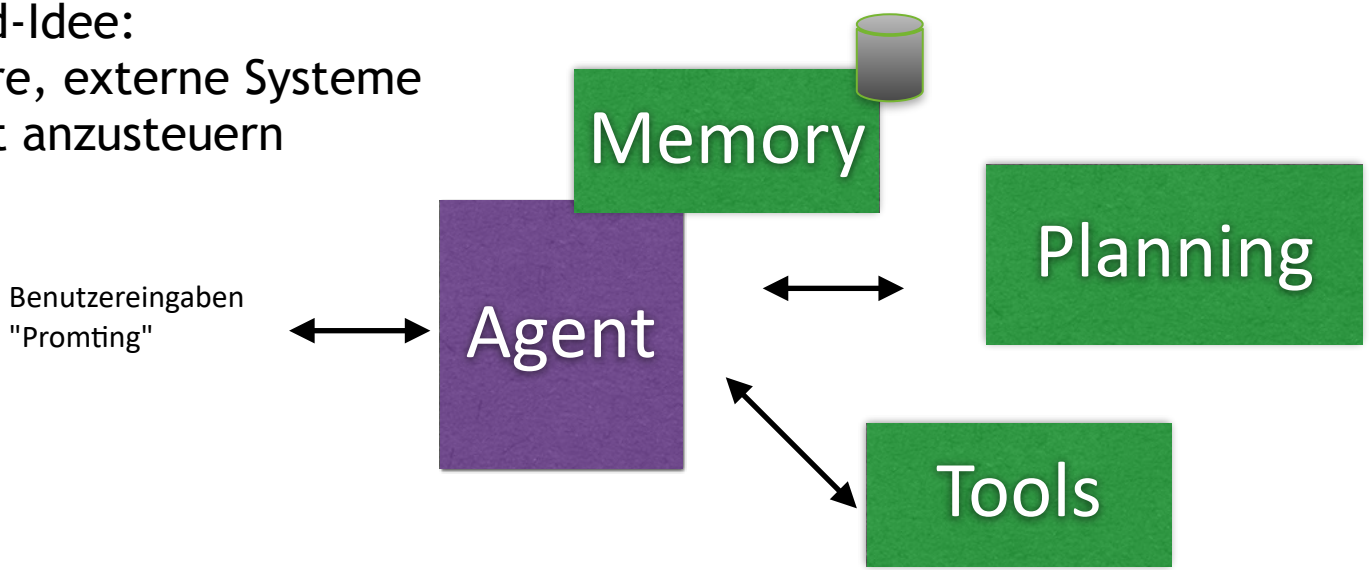
Möglicherweise ein Problem in der nahen Zukunft: Modelle trainieren sich mit zuvor selbst generierten Daten?



Aktuelle Trends / Ausblick

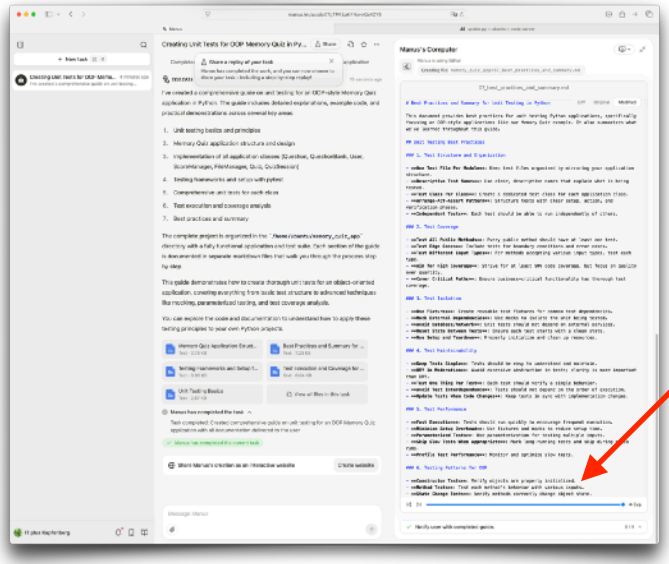
AI Agenten

Grund-Idee:
Andere, externe Systeme direkt anzusteuern



(1) Sehr kurzer **Prompt**

Show steps and extensive example to create unit tests in python for OOP style Memory Quizz application



(2) ManusAI erstellt sich einen **Plan**

(3) Schritt für Schritt Umsetzung mit **Live Vorschau**

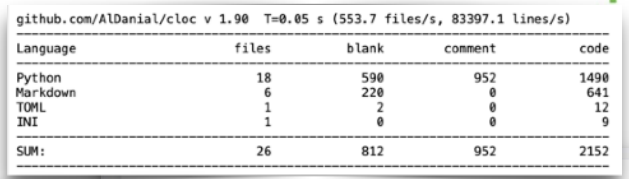
(4) Finales,
funktionierendes Projekt
in 7 bis 8 Minuten

Zirka 1500
Zeilen

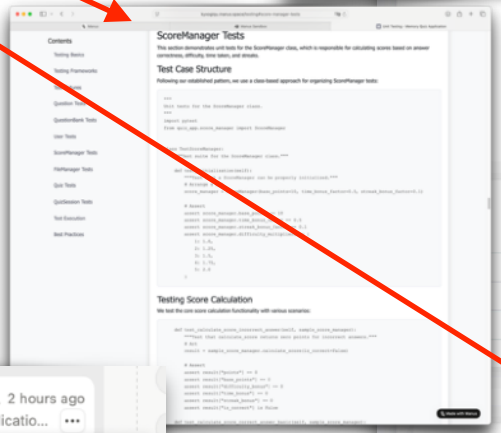
Doku

Ausprobieren:
Web version

Software Tests
durchführen



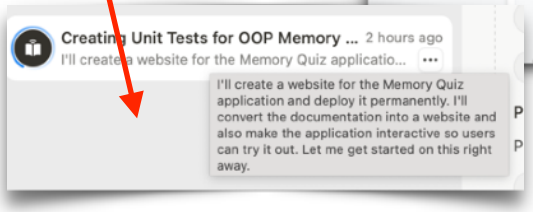
```
github.com/AlDanial/cloc v 1.90  T=0.05 s (553.7 files/s, 83397.1 lines/s)
-------------------------------------------------------------------------------
Language                     files          blank        comment           code
-------------------------------------------------------------------------------
Python                          18            590            952           1490
Markdown                         6            220              0            641
TOML                             1              2              0             12
INI                              1              0              0              9
-------------------------------------------------------------------------------
SUM:                            26            812            952           2152
-------------------------------------------------------------------------------
```



ScoreManager Tests

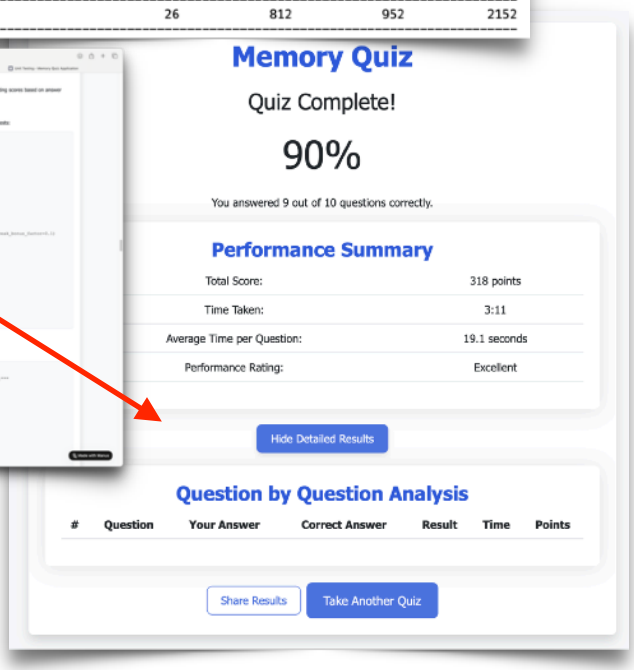
Test Case Structure

Testing Score Calculation



Creating Unit Tests for OOP Memory ... 2 hours ago
I'll create a website for the Memory Quiz applicatio...

I'll create a website for the Memory Quiz application and deploy it permanently. I'll convert the documentation into a website and also make the application interactive so users can try it out. Let me get started on this right away.



Memory Quiz

Quiz Complete!

90%

You answered 9 out of 10 questions correctly.

Performance Summary

Total Score:	318 points
Time Taken:	3:11
Average Time per Question:	19.1 seconds
Performance Rating:	Excellent

Hide Detailed Results

Question by Question Analysis

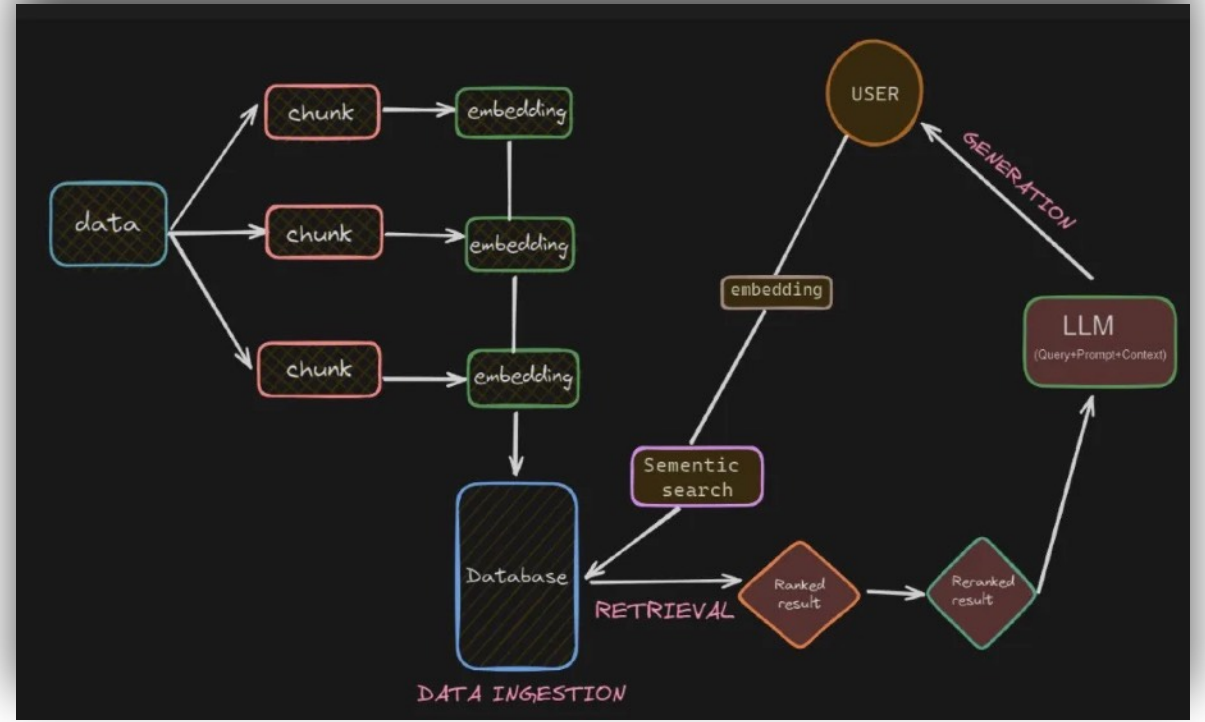
#	Question	Your Answer	Correct Answer	Result	Time	Points

Share Results | Take Another Quiz

RAG

Grund-Idee: Antworten im Kontext von eigenen (Firmen-)Daten zu erhalten

Retrieval Augmented Generation



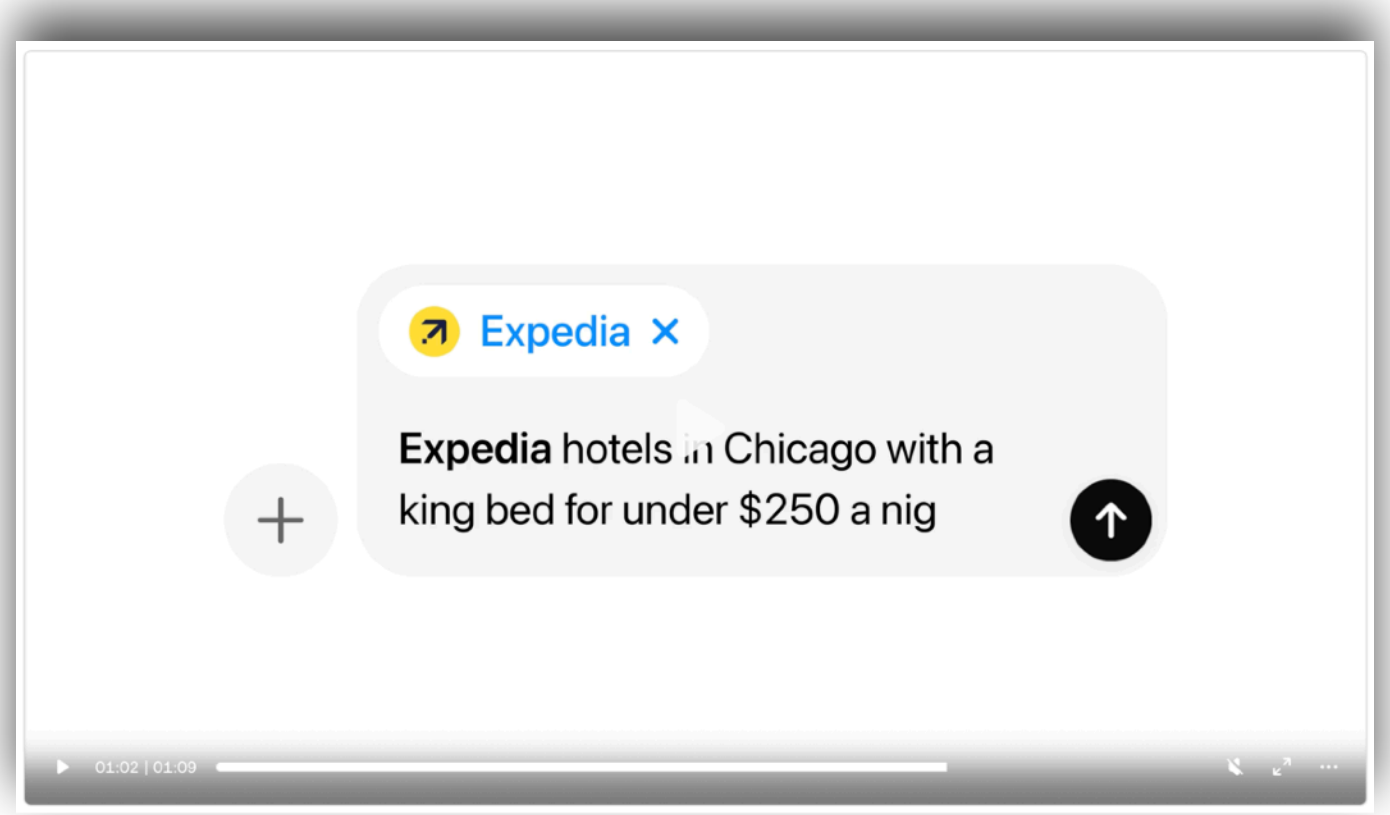
https://medium.com/@danushidk507/rag-with-llama-using-ollama-a-deep-dive-into-retrieval-augmented-generation-c58b9a1cfcd3

OpenAI Developer-Day Oktober 2025

Apps inside
= ... money, money, money, ...

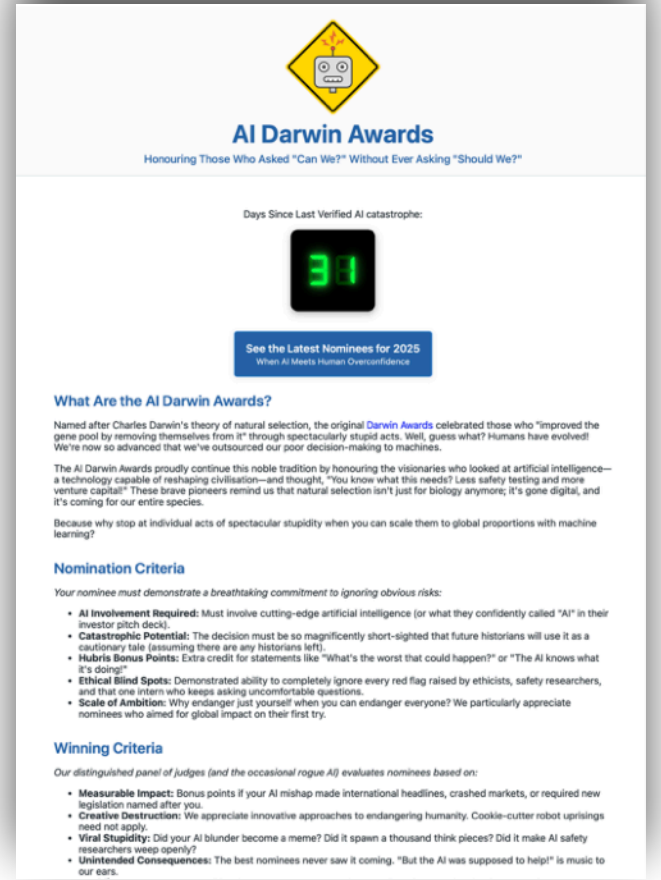
OpenAI Dev-Day

ChatGPT Hat bezahlte Services direkt integriert.



https://openai.com/index/introducing-apps-in-chatgpt/

Ai Darwin Awards



AI Darwin Awards

Honouring Those Who Asked "Can We?" Without Ever Asking "Should We?"

Days Since Last Verified AI catastrophe:

31

See the Latest Nominees for 2025
When AI Meets Human Overconfidence

What Are the AI Darwin Awards?

Named after Charles Darwin's theory of natural selection, the original Darwin Awards celebrated those who "improved the gene pool by removing themselves from it" through spectacularly stupid acts. Well, guess what? Humans have evolved! We're now so advanced that we've outsourced our poor decision-making to machines.

The AI Darwin Awards proudly continue this noble tradition by honouring the visionaries who looked at artificial intelligence—a technology capable of reshaping civilisation—and thought, "You know what this needs? Less safety testing and more venture capital!" These brave pioneers remind us that natural selection isn't just for biology anymore; it's gone digital, and it's coming for our entire species.

Because why stop at individual acts of spectacular stupidity when you can scale them to global proportions with machine learning?

Nomination Criteria

Your nominee must demonstrate a breathtaking commitment to ignoring obvious risks:

- **AI Involvement Required:** Must involve cutting-edge artificial intelligence (or what they confidently called "AI" in their investor pitch deck).
- **Catastrophic Potential:** The decision must be so magnificently short-sighted that future historians will use it as a cautionary tale (assuming there are any historians left).
- **Hubris Bonus Points:** Extra credit for statements like "What's the worst that could happen?" or "The AI knows what it's doing!"
- **Ethical Blind Spots:** Demonstrated ability to completely ignore every red flag raised by ethicists, safety researchers, and that one intern who keeps asking uncomfortable questions.
- **Scale of Ambition:** Why endanger just yourself when you can endanger everyone? We particularly appreciate nominees who aimed for global impact on their first try.

Winning Criteria

Our distinguished panel of judges (and the occasional rogue AI) evaluates nominees based on:

- **Measurable Impact:** Bonus points if your AI mishap made international headlines, crashed markets, or required new legislation named after you.
- **Creative Destruction:** We appreciate innovative approaches to endangering humanity. Cookie-cutter robot uprisings need not apply.
- **Viral Stupidity:** Did your AI blunder become a meme? Did it spawn a thousand think pieces? Did it make AI safety researchers weep openly?
- **Unintended Consequences:** The best nominees never saw it coming. "But the AI was supposed to help!" is music to our ears.

https://aidarwinawards.org/

ENDE

https://aibreakthroughawards.com/2024-winners/