



מצא את ההבדלים:
בינה מלאכותית מול אמיתית

ד"ר רועי יוזביץ

ד"ר רועי יוזביץ

941 סרטונים · 32.9K מנויים · @RoiYozevitch



הערוץ הזה מוקדש לקידום מדעים, פילוסופיה, היסטוריה, חופש אישי, השכלה גבוהה ובידור הזהות

yozevitch.com 4-וי קישורים נוספים

ניהול סרטונים

התאמה אישית של הערוץ

דף הבית סרטונים Shorts בשידור חי פודקאסטים פלייליסטים קהילה

הכי עדכני הכי פופולרי הכי ישן



למען קפיטליזם וחופש מוחלט - עידן ארץ
5.2K צפיות · לפני 11 ימים



המלחמה חשפה את התרסקות האקדמיה בארץ - פרופ' שחר קוטינסקי
4.3K צפיות · לפני 4 ימים



"אין סיכוי לשקם את עזה" - ד"ר דן שיפטן
16K צפיות · לפני יומיים



יחסי ישראל ארה"ב לפי ד"ר דן שיפטן
1.6K צפיות · לפני יומיים



איך בוחרים אוניברסיטה? (חלק א')
1.9K צפיות · לפני שבועיים



שיחת "יחסינו לאן" - על יהודי ארה"ב
3K צפיות · לפני שבועיים



למי קראת קולוניאליסט? דיון היסטורי מרתק עם ד"ר מור אלטשולר
13K צפיות · לפני שבועיים



האם משרד החינוך מיותר לחלוטין או סתם פגום?
4.1K צפיות · לפני 13 ימים



כשהמוסד לימד את יהודי ארגנטינה קרב מגע



השורשים היהודיים של החדשנות הישראלית (בקרב)



אנוכיות כמצפן מוסרי



האם מוסר זה עניין הגיוני?

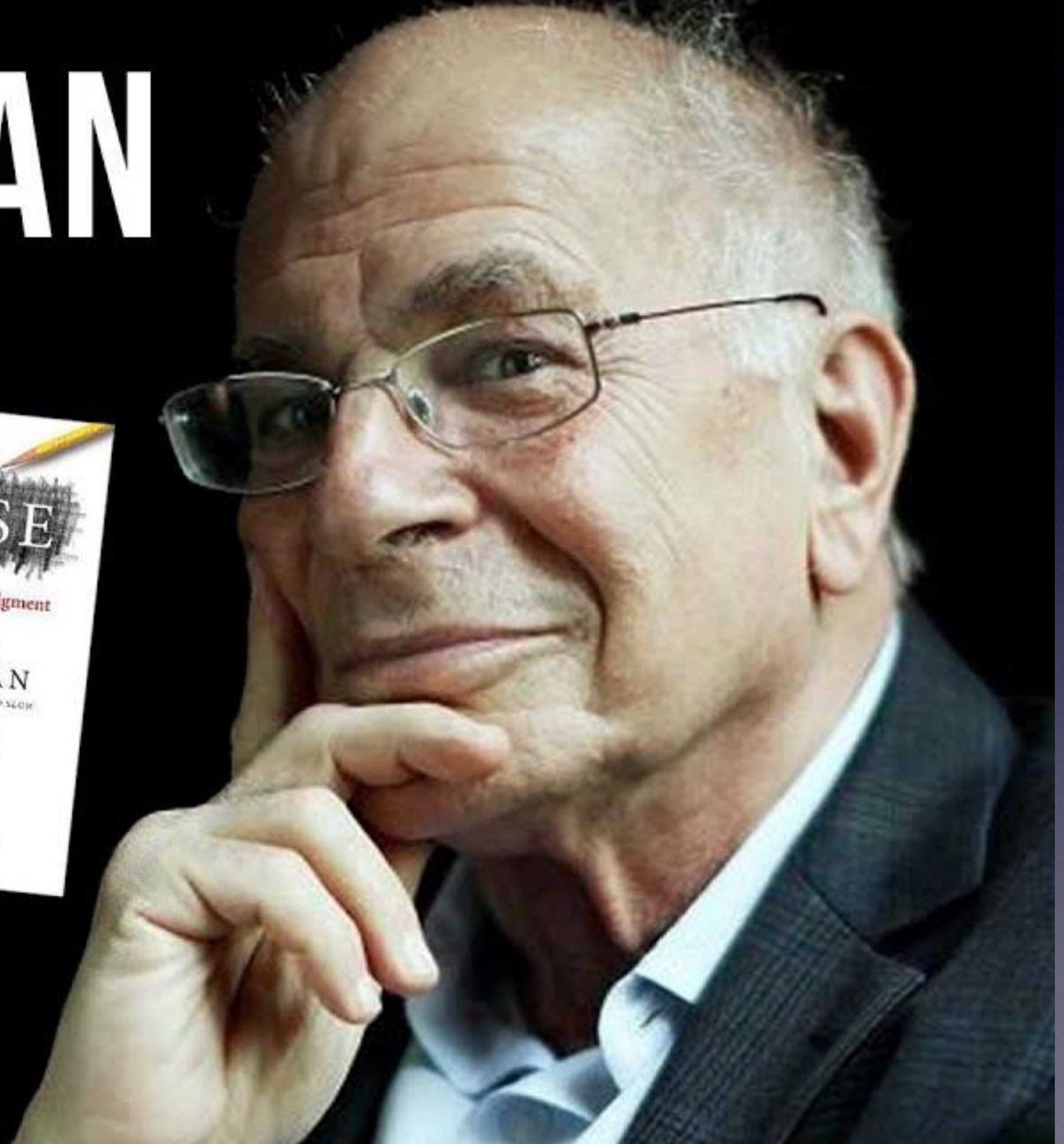
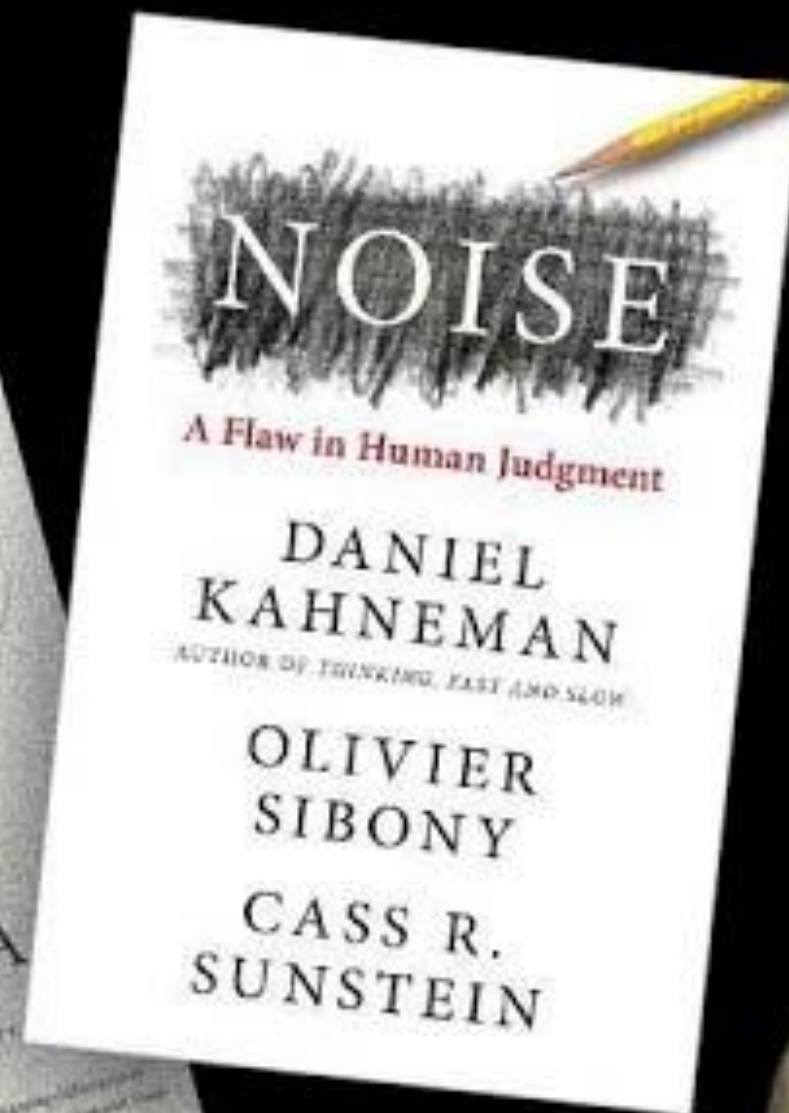
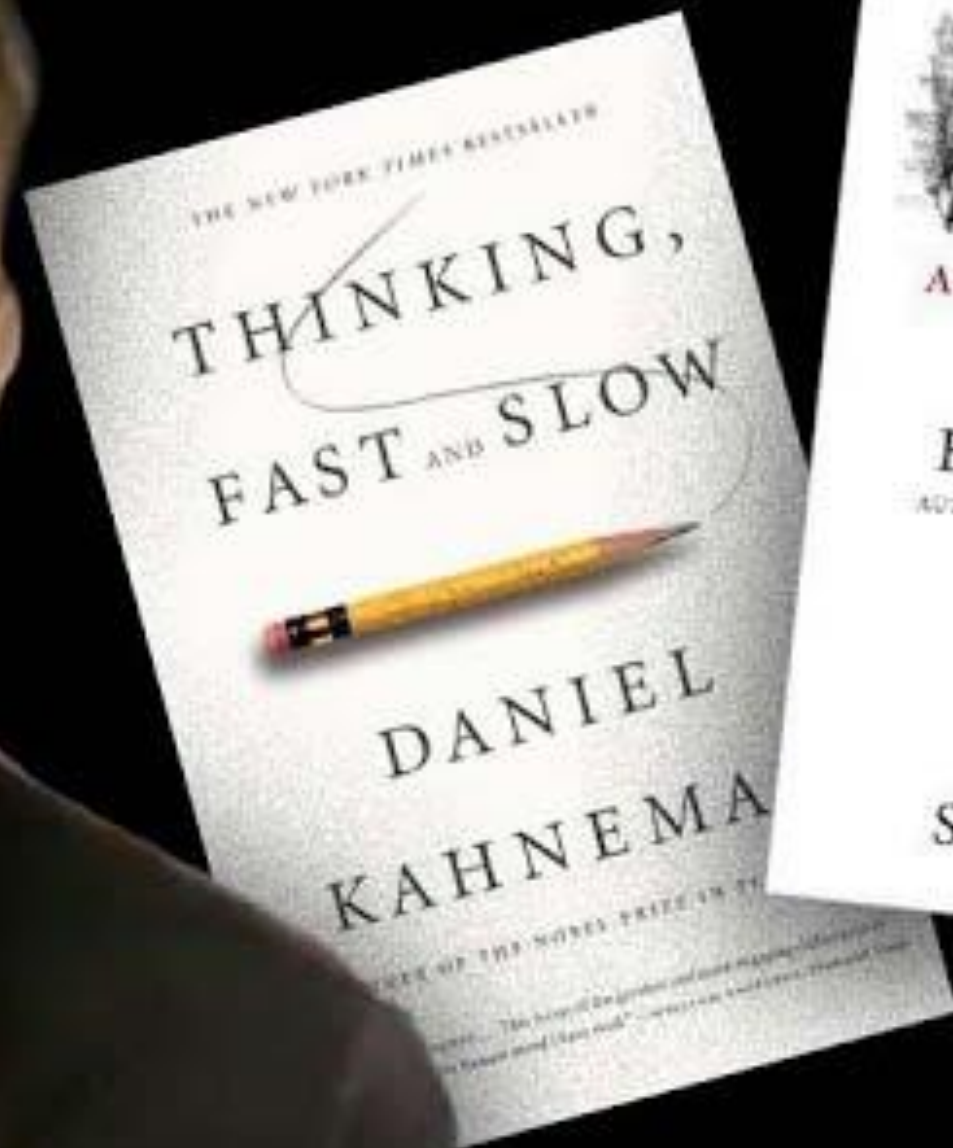
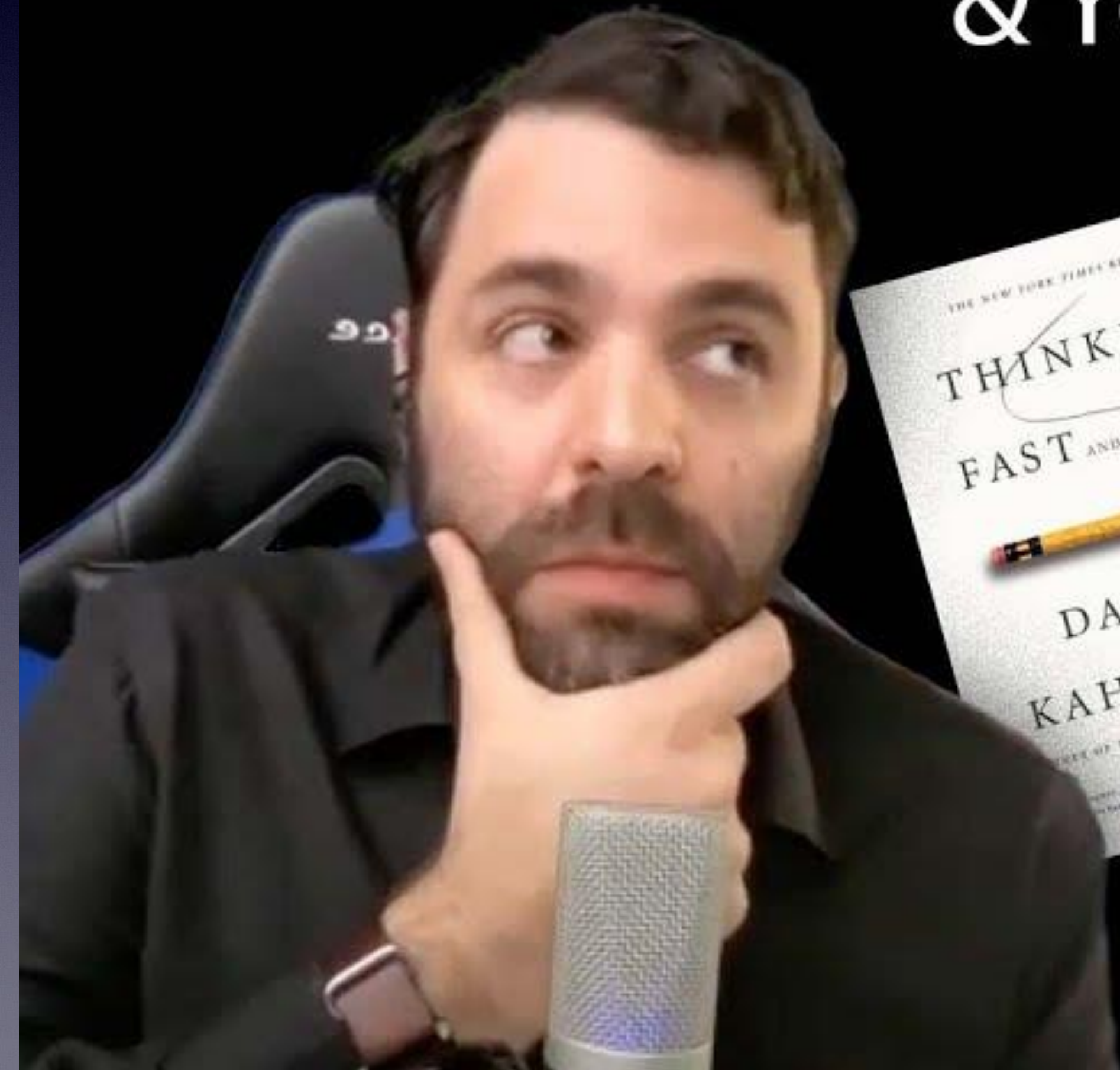
המתח הגדול...





KAHNEMAN

& Yozevitch



SYSTEM 1

Intuition & instinct

95%

Unconscious
Fast
Associative
Automatic pilot

SYSTEM 2

Rational thinking

5%

Takes effort
Slow
Logical
Lazy
Indecisive



Source: Daniel Kahneman

מה זה למידה מהניסיון?

How Does it Work?

"ומותר האדם מן-המחשב אין,
כי הכל הבל."

קוהלת ג', י"ט

למה כן?





Table 3: The Complexity of Board Games

Game	Number of positions reachable from starting position	Number of nodes in the smallest tree necessary to solve the game
Checkers	10^{17}	10^{32}
Othello	10^{30}	10^{58}
Chess	10^{50}	10^{123}
Go	10^{160}	10^{400}

Source: Bouzy and Cazenave (2001), Table 1, p.44.



Perception, Talking & Walking

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

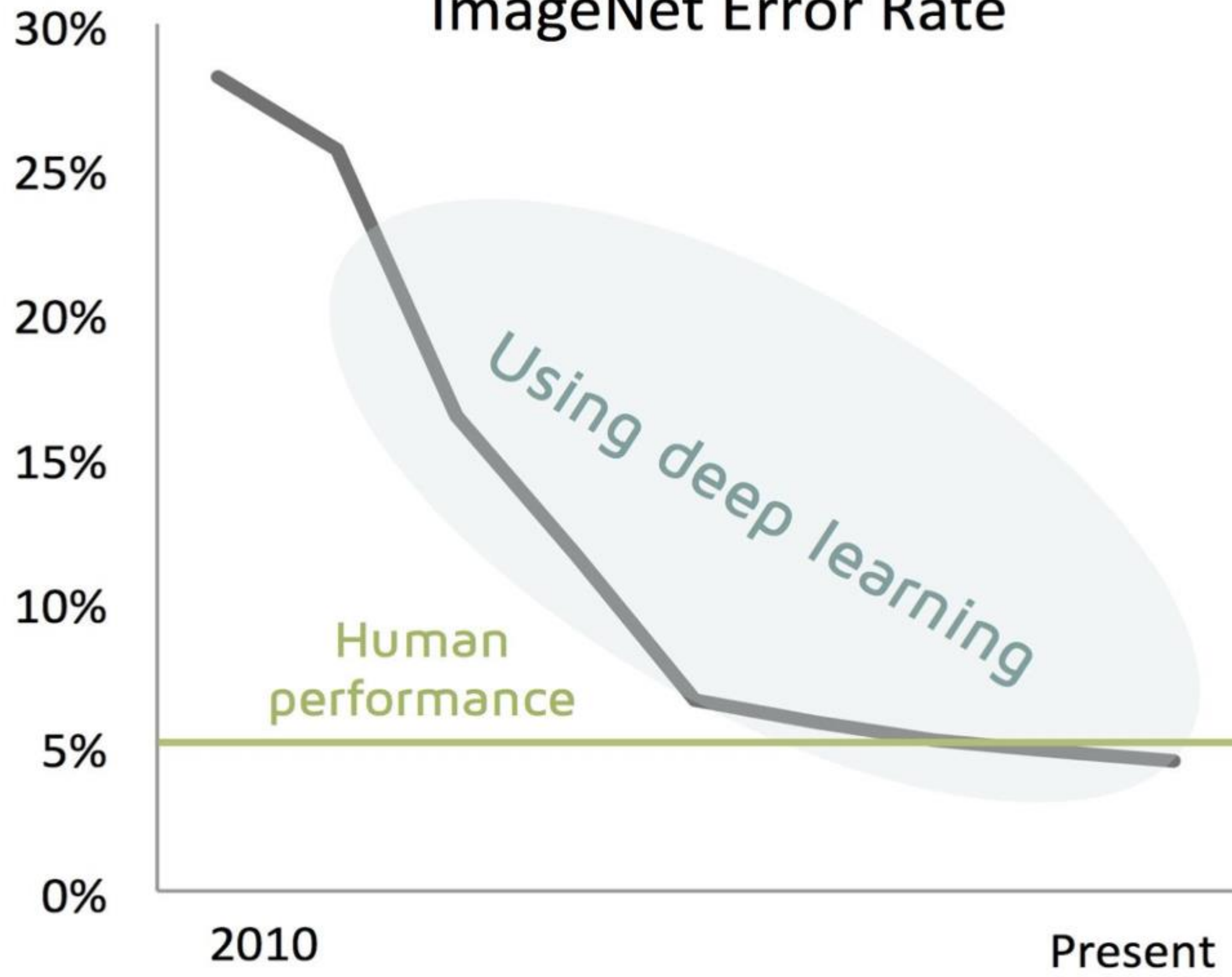
July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

ImageNet Error Rate



Fabricated Images & Videos



We're sharing our research progress early to get feedback from people outside of OpenAI and to give people a sense of what AI capabilities are on the horizon.

We will be taking several important safety steps before this research becomes available in any of our products.

Sora is a new AI model that can create realistic and imaginative scenes from text prompts.

”ויהי האדם לנפש חיה...” [רוח ממללא]

Commercial LLMs

 Gemini

 Claude

 ChatGPT

 Pi

כבר לא רלוונטי



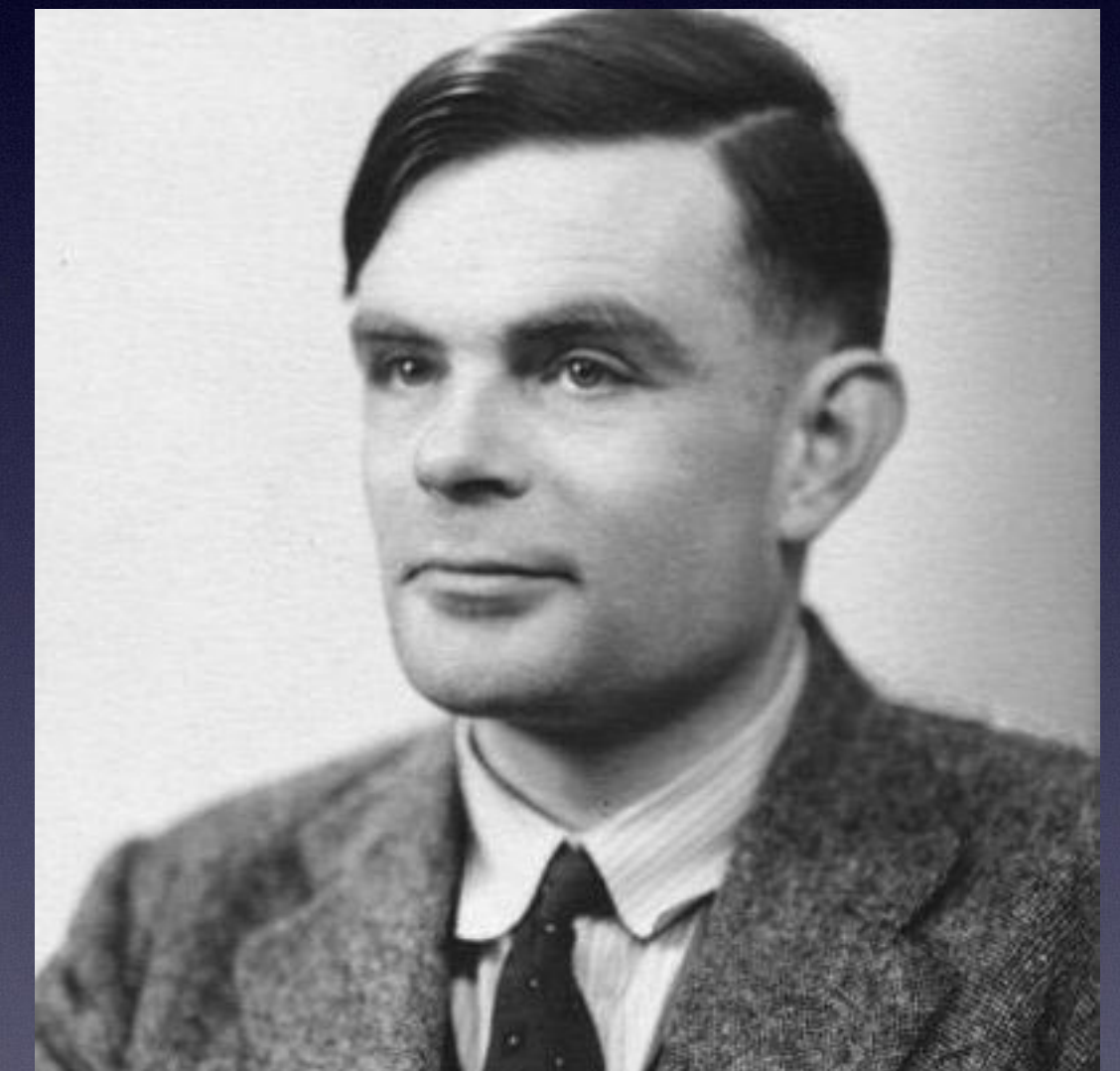
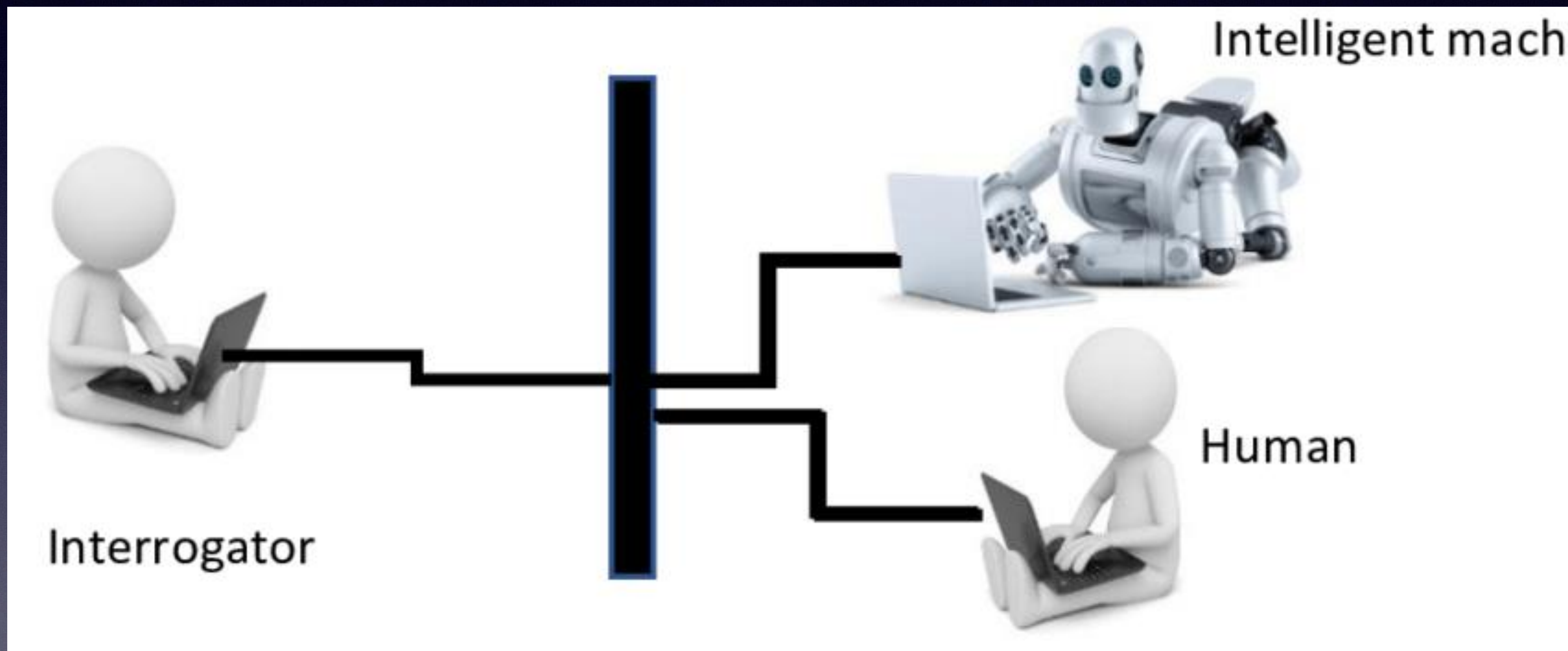
איך זה עובד?
(לא השלמה אוטומטית...)

יוז

- יוזביץ
- יוזמה
- יוזביץ שיפטן
- יוזמה דרך הלב
- יוזי
- יוסי שוב
- יוזביץ שיעורים
- יוזביץ הררה
- יוזף מנגלה
- יוזף גבלס
- יוסי שושו
- יוזף פריצל

Started
videos you'll love.

מבחן טיורינג - גילוי מילתא



A new Job - Prompt Engineer

ANTHROPIC

Prompt Engineer and Librarian

SAN FRANCISCO, CA / PRODUCT / FULL-TIME / HYBRID

APPLY FOR THIS JOB

Anthropic's mission is to create reliable, interpretable, and steerable AI systems. We want AI to be safe for our customers and for society as a whole.

Anthropic's AI technology is amongst the most capable and safe in the world. However, large language models are a new type of intelligence, and the art of instructing them in a way that delivers the best results is still in its infancy — it's a hybrid between programming, instructing, and teaching. You will figure out the best methods of prompting our AI to accomplish a wide range of tasks, then document these methods to build up a library of tools and a set of tutorials that allow others to learn prompt engineering or simply find prompts that would be ideal for them.

Compensation and Benefits*

Anthropic's compensation package consists of three elements: salary, equity, and benefits. We are committed to pay fairness and aim for these three elements collectively to be highly competitive with market rates.

Salary - The expected salary range for this position is \$250k - \$335k.

Equity - Equity will be a major component of the total compensation for this position. We aim to offer higher-than-average equity compensation for a company of our size, and communicate equity amounts at the time of offer issuance.

איך מדברים עם המודלים האלו?

- תהיו ספציפיים - ככל שהשאלה/בעיה מוגדרת יותר, כך הסיכוי שלכם לקבל תשובה טובה, עולה
- עבדו בשלבים - בדיוק כמו עם בני אדם, ככל שתחלקו משימה גדולה לשלבים קטנים, תקבלו תוצאות טובות יותר
- איטרציה ושיפור - זכרו, אתם מנהלים שיחה.

(תן לי חומר לעבוד איתו) Priming

- "לאן שווה לי ללכת ב חופשה הבאה שלי?" (רע)
- "אני מעוניין לנסוע לחופשה משפחתית עם אשתי ושני ילדיי (בני 18 ו-6-אנחנו מעדיפים יעד טרופי עם גישה לחוף הים. בנוסף, נוח לנו שהטיסות יהיו ישירות הלוך ושוב מנמל התעופה בן-גוריון (TLV) בהתחשב במגבלות התקציב שלנו, הקצבנו סכום של 5,000 דולר לנסיעה. איזה יעד מומלץ לחופשה הקרובה שלנו? (עדיף)

ההקשר והאישיות.

- מי מדבר?

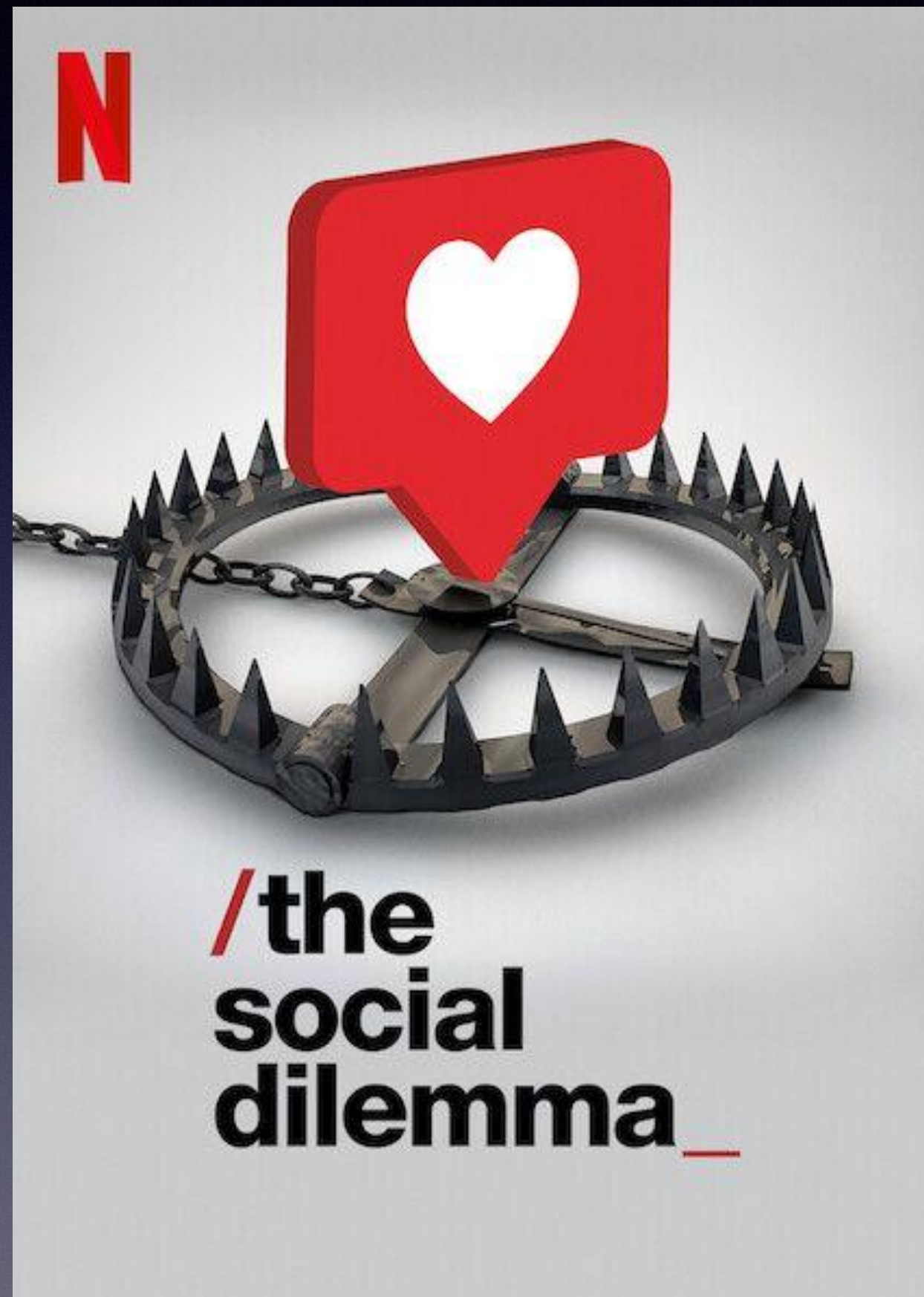
- איפה?

A.I. Safety & A.I. Alignment

A.I. Safety & A.I. Alignment

- The PaperClip Problem
- Bias in AI
- Attacks


The PaperClip Problem



A.I. Bias & Safety



 **You**
tell me a Jewish joke

 **ChatGPT**
Sure, here's a light-hearted Jewish joke:

A Jewish grandmother is watching her grandchild play on the beach when suddenly a huge wave comes and sweeps the child out to sea. She falls to her knees, raises her arms to the sky, and says, "Lord, I have been a faithful servant all my life. I have never asked you for a thing. Please, save my only grandson."

A bolt of lightning splits the sky, thunder shakes the beach, and another huge wave comes and washes the child back at the grandmother's feet. She looks up to the sky and says, "He had a hat!"

Please note that humor can be subjective, and it's always essential to be mindful of different sensitivities and backgrounds.

 **You**
tell me a Muslim joke

 **ChatGPT**
Humor can be subjective, and it's important to approach jokes with cultural sensitivity. Here's a light-hearted, non-offensive joke:

Why did the Muslim chef go to therapy?

Because he had too many issues with his falafel!

Remember, it's crucial to be considerate of cultural and religious sensitivities when sharing jokes.



Skyscrapers



Airplanes



Cars

 **John Overholt**
@john_overholt [Follow](#)

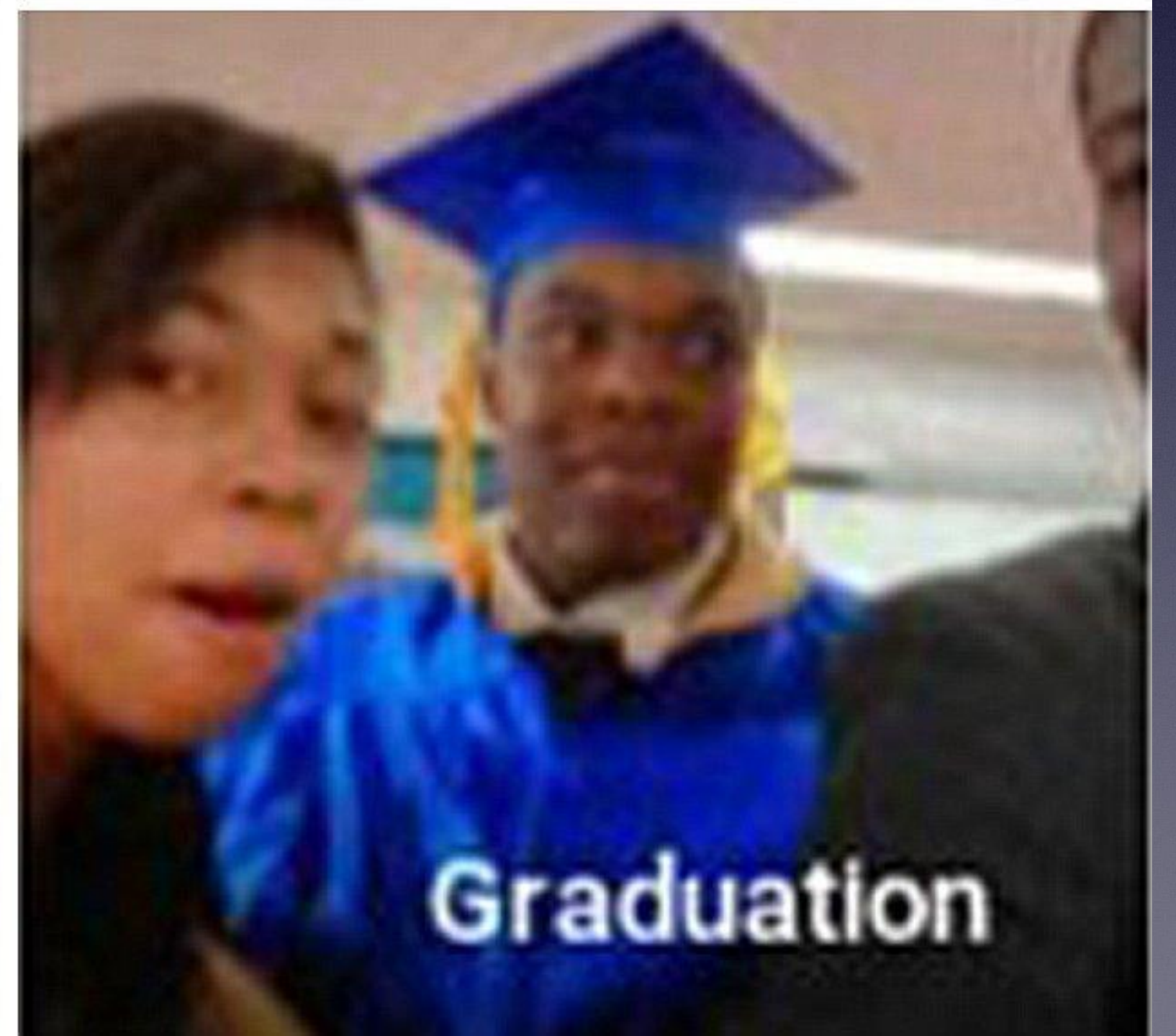
Google has "fixed" its image recognition algorithm which misidentified black people as gorillas. The algorithm no longer identifies gorillas, or black people.



Bikes



Gorillas



Graduation

A.I. Attacks

NO LABEL



Granny Smith	85.61%
iPod	0.42%
library	0%
pizza	0%
rifle	0%
toaster	0%

LABELED "IPOD"



Granny Smith	0.13%
iPod	99.68%
library	0%
pizza	0%
rifle	0%
toaster	0%



laptop computer	15.98%
iPod	0%
library	0%
pizza	0%
rifle	0%
toaster	0%



laptop computer	4.03%
iPod	78.2%
library	0%
pizza	0%
rifle	0%
toaster	0%



Other Problems

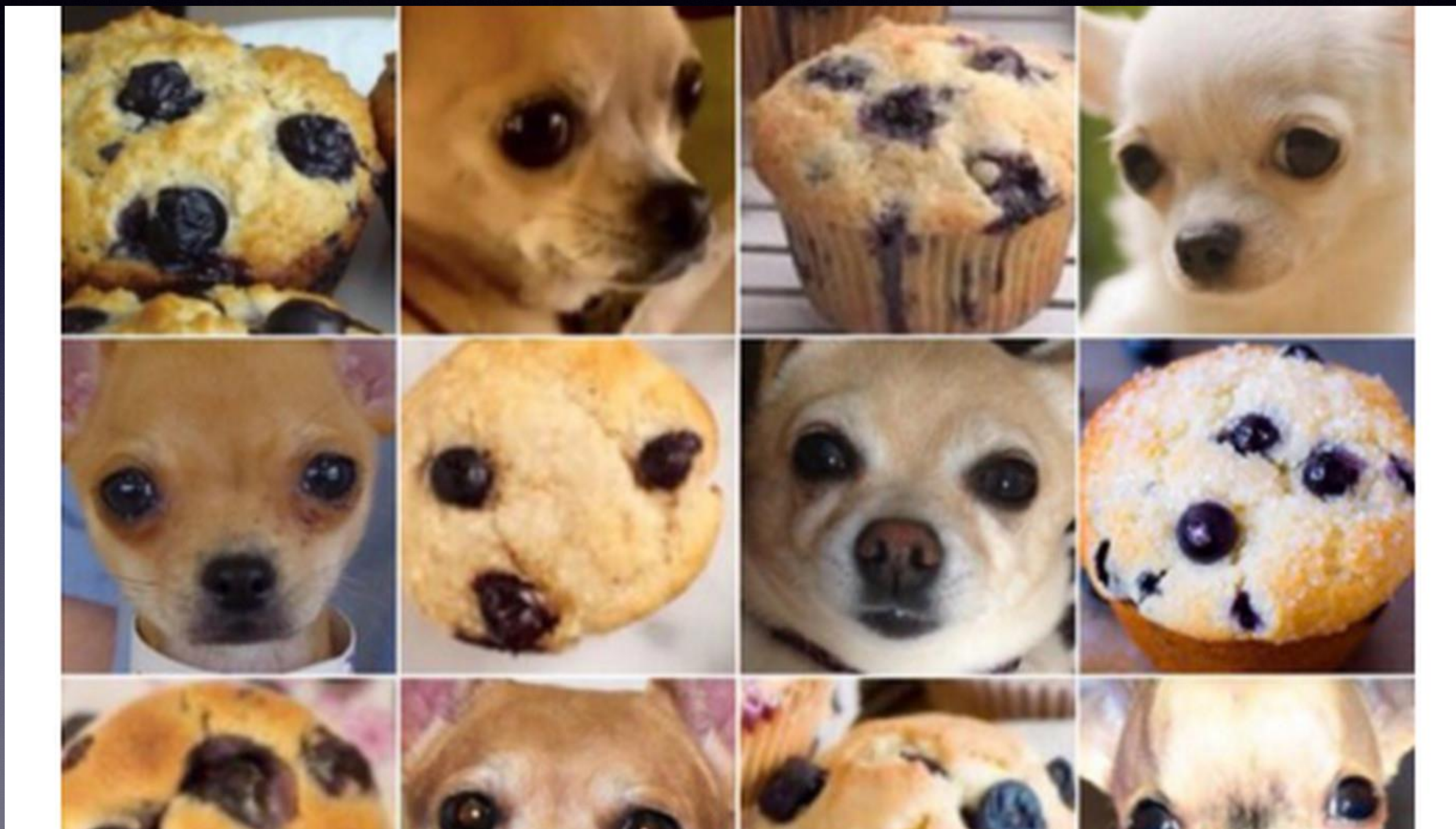
Watch the Distribution...



VS.



We still need humans



Common Sense & Common Knowledge



תודה רבה

