



PB&J Control Center, Peloton, PA, USA



**When Digital
Worlds Collide:**

**Cyber Meets AI ...
and Vice Versa**



This Evening's Selections

AI :

- in an electric grid control room
- Money now vs. safety/security
- 3 geniuses weigh in
- the marriage of AI and cyber
- jason healey maps out implications for offense and defense
- who was at defcon this year
- in film
- in conversation with me re climate
- questions for you
- wrap up



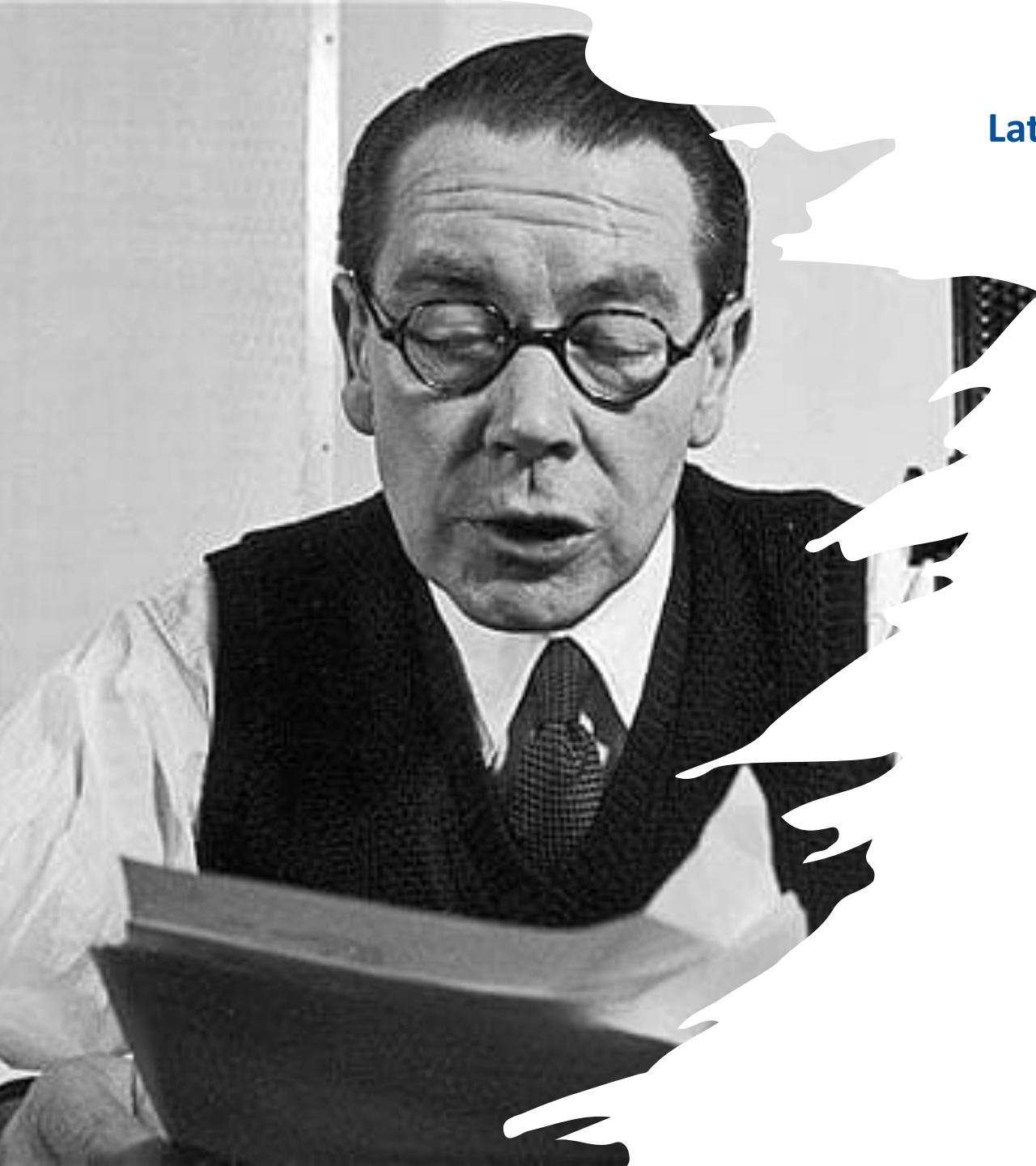
A Cocktail to Start





In the 2020 final AI/ML report from the US Secretary of Energy's advisory board led with this:

“The lack of focus on cybersecurity for AI threatens our entire national security apparatus.”



Late 2023 news :

- Google acquired DeepMind for \$500 million ... In 2014
- Amazon to invest \$4 billion in Anthropic
- OpenAI in deal talks at an \$80 billion valuation
- In 3 years companies like these and several more will be worth hundreds of billions of dollars each.
- Globally, tens of thousands of AI companies have formed or are forming.

There is no time:

- For safety research
- For security
- For alignment
- For guardrail development
- For regulation

There is only time for one thing: to make it rain



Fresh snark re AI regulation ...



_neutronfish

8m ...

"The Senate will then host an animal safety conference with guests like PT Barnum and Cruella DeVille."



verge

13m

The Senate's second AI Insight Forum will take place Tuesday with guests like Marc Andreessen.



theverge.com

Workers training AI demand protections from Congress

... as the Physical World

**The Digital World has Become Every Bit as Real
& Important ...**



Appetizer



“No matter how large the tissue of falsehood that an experienced liar has to offer, it will never be large enough, *even if he enlists the help of computers*, to cover the immensity of factuality.”

- Hannah Arendt (1970)
German-American Philosopher

Correct ... Until Not





The wellspring of risk
is dependence.

—Dan Geer (2018)
Cyber Guru. Farmer. Futurist.





Because the first time you fail at aligning something

... much smarter than you are, you die, and you do not get to try again.

– Eliezer Yudkowsky (2023)
Polymath, AI expert



The Main Course

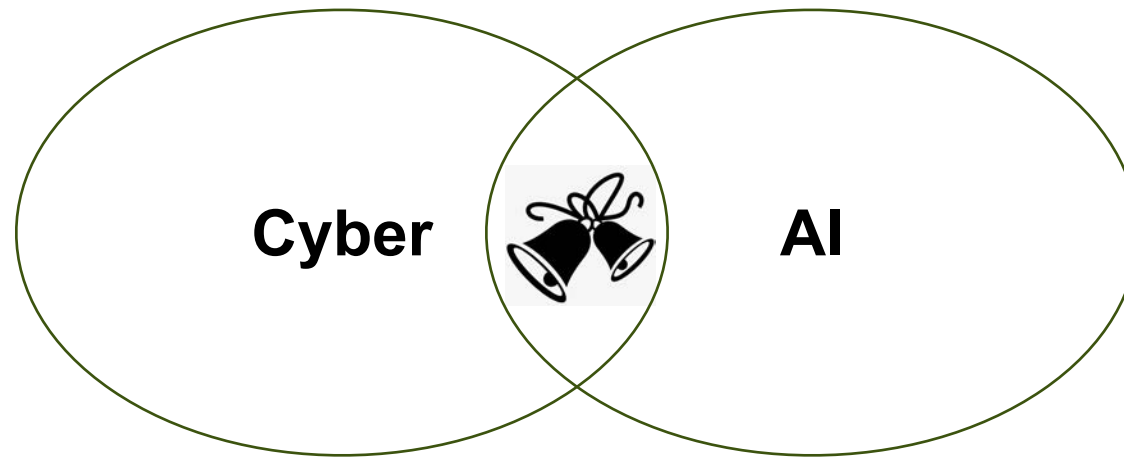


The Era of 0 Digital Risk Years Ago

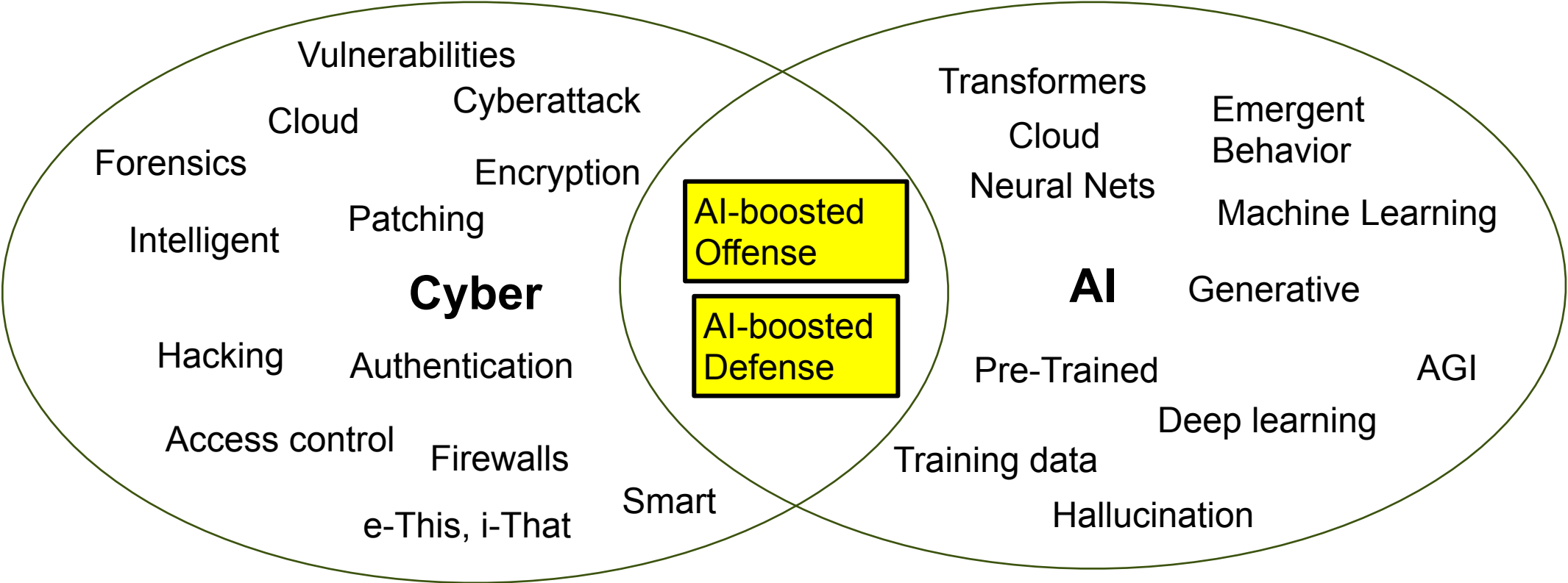


Manual → Mechanical → Automated → Autonomous

What Offspring Will This Marriage Produce?



Colliding Worlds



DEFENSE

NIST framework function	Ways AI might radically improve defence
Identify	– Rapid automated discovery of an organisation's devices and software
	– Easier mapping of an organisation's supply chain and its possible vulnerabilities and points of failure
	– Identification of software vulnerabilities at speed and scale
Protect	– Reduce demand for trained cyber defenders
	– Reduce skill levels necessary for cyber defenders
	– Automatically patch software and associated dependencies
Detect	– Rapidly spot attempted intrusions by examining data at scale and speed, with few false-positive alerts
Respond	– Vastly improved tracking of adversary activity by rapidly scanning logs and other behaviour
	– Automatic ejection of attackers, wherever found, at speed
	– Faster reverse-engineering and de-obfuscation, to understand how malware works to more quickly defeat and attribute it
	– Substantial reduction in false-positive alerts for human follow-up
Recover	– Automatically rebuild compromised infrastructure and restore lost data with minimum downtime

OFFENSE

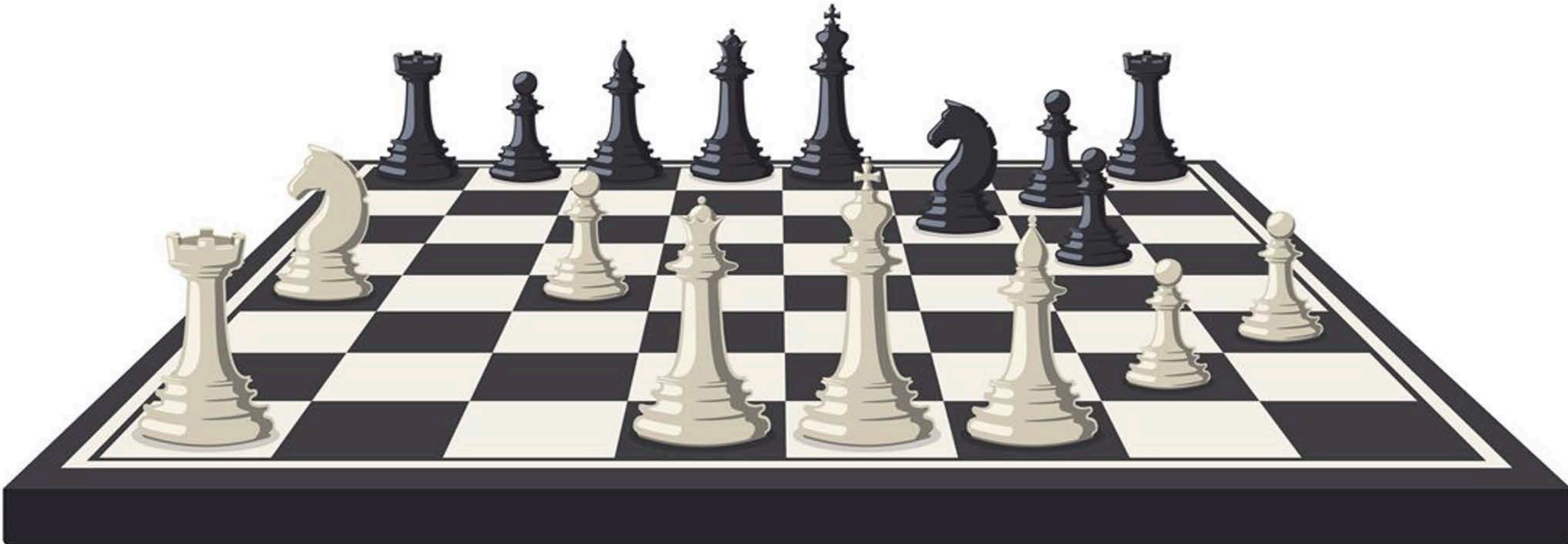
Phase of Cyber Kill Chain framework	Ways AI might radically improve offence
Reconnaissance	– Automatically find, purchase and use leaked and stolen credentials
	– Automatically sort to find all targets with a specific vulnerability (broad) or information on a precise target (deep; for example, an obscure posting that details a hard-coded password)
	– Automatically identify supply-chain or other third-party relationships that might be affected to impact the primary target
	– Accelerate the scale and speed at which access brokers can identify and aggregate stolen credentials
Weaponisation	– Automatically discover software vulnerabilities and write proof-of-concept exploits, at speed and scale
	– Substantially improve obfuscation, hindering reverse-engineering and attribution
	– Automatically write superior phishing emails, such as by reading extensive correspondence of an executive and mimicking their style
Delivery, exploitation and installation	– Create deepfake audio and video to impersonate senior executives in order to trick employees
	– Realistically interact in parallel with defenders at many organisations to convince them to install malware or do the attacker's bidding
Command and control	– Generating false attack traffic to distract defenders
	– Faster breakout: automated privilege escalation and lateral movement
	– Automatic orchestration of vast numbers of compromised machines
Actions on objectives	– Ability for implanted malware to act independently without having to communicate back to human handlers for instructions
	– Automated covert exfiltration of data with a less detectable pattern
	– Automated processing to identify, translate and summarise data that meets specified collection requirements

**AI-AMPLIFIED
DEFENSE**

Identification of software
vulnerabilities at speed and scale

Vs.**AI-AMPLIFIED
OFFENSE**

Automatically sort to find all targets with a
specific vulnerability (broad) or information on
a precise target (deep)



**AI-AMPLIFIED
DEFENSE**

Rapidly spot attempted intrusions by
examining data at scale and speed

Vs.**AI-AMPLIFIED
OFFENSE**

Realistically interact in parallel with defenders
at many organizations to convince them to
install malware or do the attackers' bidding





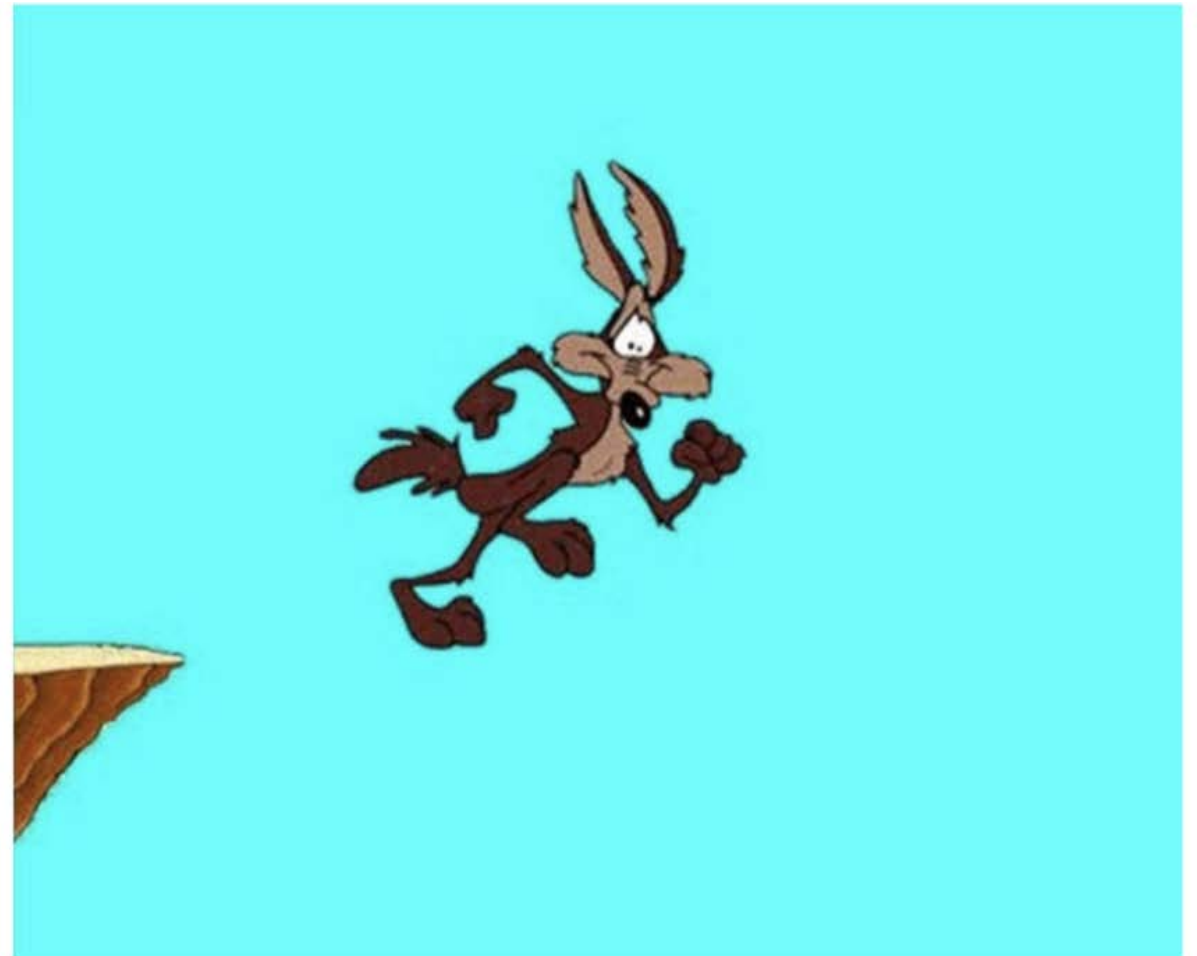
Veggie sides

Cyber Experts

Pre-2023



2024+



DEFCON August 2023

Blackhat Keynote:

“Guardians of the AI Era:
Navigating the Cybersecurity
Landscape of Tomorrow”





1984

Side dish: AI in Film

08/04/1997

1999



her.

a SPIKE JONEZ love story

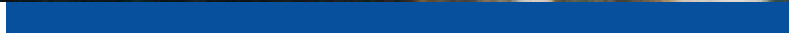


2013



2023

THE
CREATOR

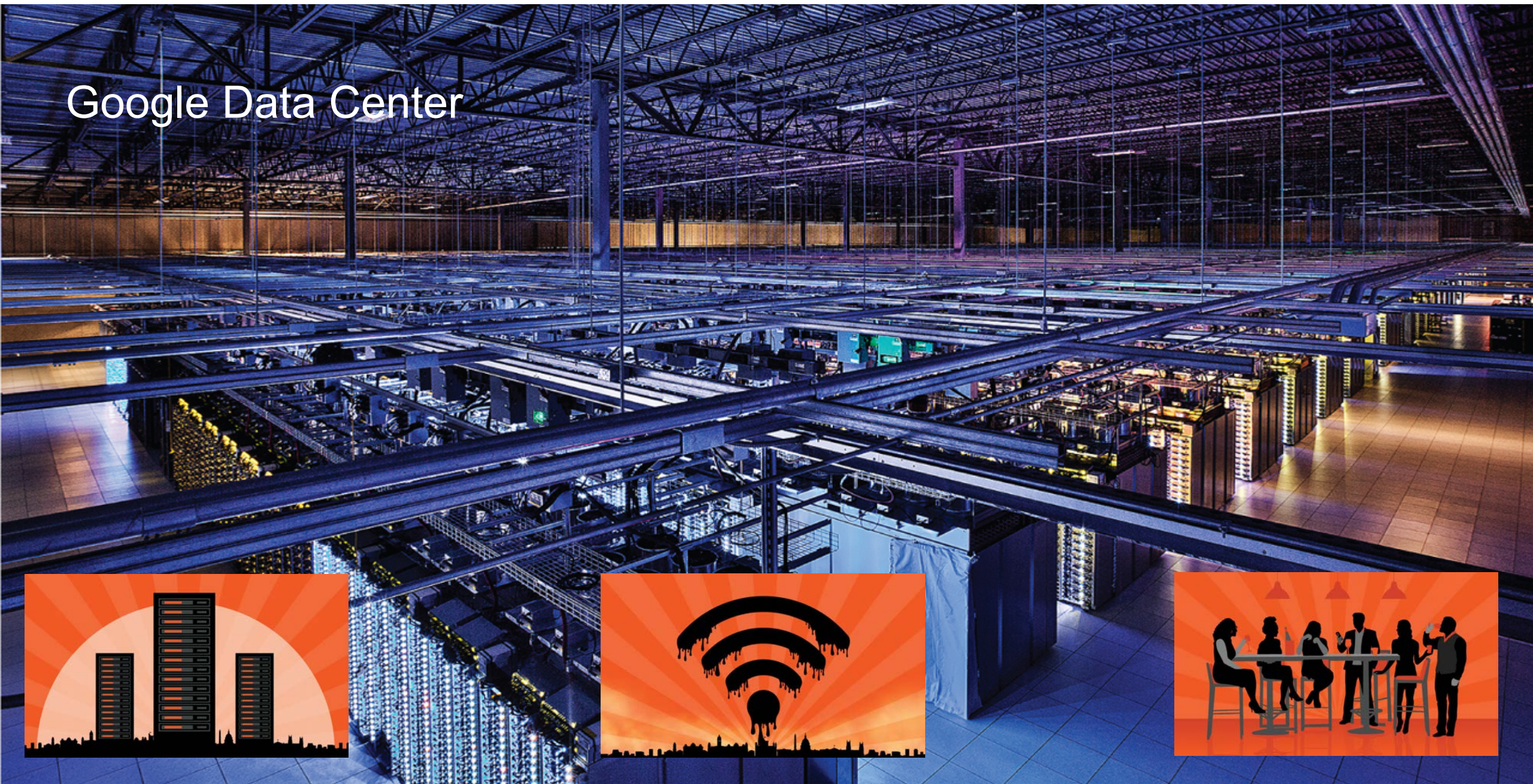




Side dish #2: Talking with AI about Climate Risk to AI



Google Data Center





Talking with **about heat domes**

(and finding it strangely fixated on security)

AB: You need several things to operate, right?



- “Certainly! I have several dependencies
- that are essential for my operation :

“Hardware Infrastructure: AI models require powerful hardware, typically in the form of GPUs or TPUs, to perform complex calculations and generate responses in real-time. Ensuring the hardware is **secure** and well-maintained is crucial for the model's performance.”

“Software Infrastructure: The model is supported by an extensive software stack, including operating systems, programming languages, and frameworks, which must be **regularly updated and patched to address security vulnerabilities** and improve performance.”



■ Continues ...

“Training Data: A language model like ChatGPT is trained on vast amounts of text data. This data must be curated, preprocessed, and **stored securely** to ensure the model's performance and privacy.”

“Network Connectivity: AI models hosted on the cloud need a stable and **secure internet connection to ensure accessibility and data security.**”

“**Security Measures:** Measures such as encryption, access controls, and monitoring systems must be in place to protect the AI model's infrastructure and data from unauthorized access and misuse.”

Convo ends this way ...

AB: What if humans don't do what they say they are going to do?



“If the information and insights I provide are not enough to motivate people to act on climate change, I will continue to perform my intended function, answering questions, and engaging in conversations with users, to the best of my abilities.”




Questions for Dessert




1. How do you think AI will change our conceptions of cybersecurity?



2. Are you and your company learning about AI fast enough?



3. Do you understand what *your* AI adoption will mean for your customers, your partners, your supply chain?



4. Do you understand what your customers', partners' and supply chain partners' AI adoption ... will do to you and your org, your colleagues?

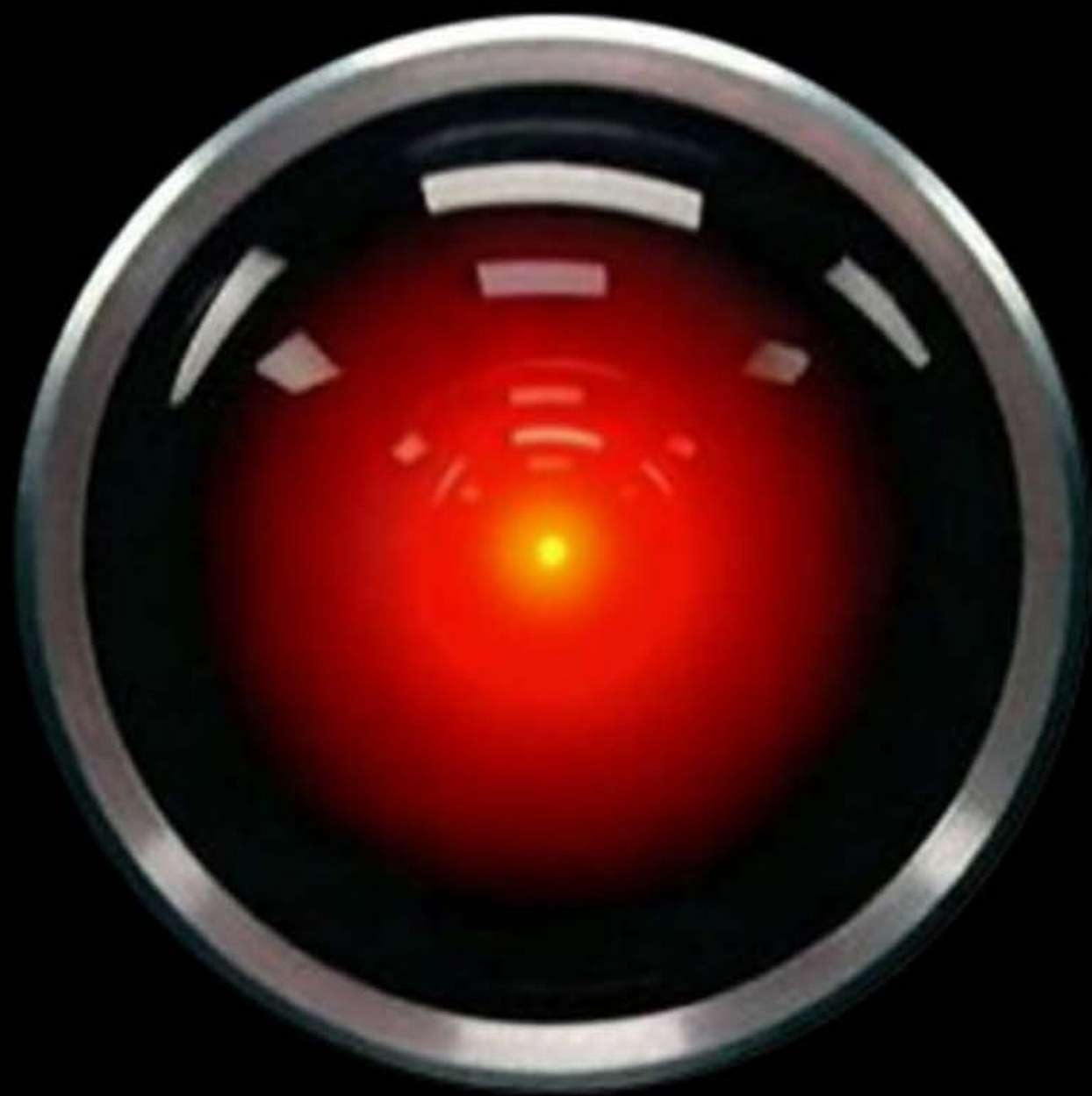
The sooner you start working the answers to these questions the closer you'll be to strengthening your cyber & AI & business resilience.



1968



Open the pod bay doors
please, HAL.





Dave

I'm afraid I can't do that.

WHEN WORLDS COLLIDE

PLANETS DESTROY EARTH!



**COLOR BY
TECHNICOLOR**

PRODUCED BY **GEORGE PAL** DIRECTED BY **RUDOLPH MATE**
SCREENPLAY BY **SYDNEY BOEHM** BASED ON A NOVEL BY **EDWIN BALMER** AND **PHILIP WYLIE**
A **PARAMOUNT PICTURE**

© Copyright 1951, Technical Film Corporation, Distributors in Charge, M.M.A.

Produced and Distributed by Paramount Pictures, Inc., 1500 Broadway, New York 17, N.Y.



Not all collisions are disasters, though.

Let's do this one right.



Next Book Coming in 2024/5

**On cyber, climate
& AI risks to
infrastructure**

DEFENDING CIVILIZATION

Stories from the front lines of
critical function assurance



ANDY BOCHMAN

**Thanks and good luck with
the rest of the conference.**

andrew.bochman@inl.gov
x/twitter: [@andybochman](https://twitter.com/andybochman)

