Robots, AI, War, Security

AI 109

Spring 2026

Plan for Today

- Some details about the final
- Data centers and resource usage
- News around AI and robots
- Finishing up the discussion of military robots
- Some thoughts about AI cybersecurity & securing yourself

Final Format

- Saturday, May 9th. 10:15 a.m.-12:15 p.m. This room.
  - Yes I know this is an awful time. I didn't pick it.
- No written exam
- Presentation instead
  - 7 minutes
  - You will use the slides you submit on Friday, so make them good
  - Speaker order determined randomly
  - I'll make a single deck with all slides to maximize efficiency.
- Grading
  - You will write evaluations of other students' presentations
  - You will be graded on both your presentation and your evaluations.
- Let Dr. Kelley know if you have any food allergies – he plans to bring donuts.

Interesting News from Virginia

(Bloomberg) -- Compass Datacenters is pulling out of a yearslong effort to build a key part of 2,100-acre data center corridor in Northern Virginia after the development faced intense pushback from local residents.

The Brookfield Asset Management-backed data center company has spent years trying to secure Prince William County's blessings to develop more than 800 acres as part of the project. After sinking tens of millions of dollars into the effort, the firm decided that public opposition and state lawmakers' growing resistance to providing tax breaks created too many roadblocks, according to people familiar with the matter, who asked not to be named discussing non-public information.

https://x.com/TheStalwart/status/2049461978167423472?s=20

Data Centers and Resources

- This probably not a real issue.
- Thermal design power (TDP)
  - Maximum amount of heat that a computer component (CPU, GPU, or system on chip) can generate and that its cooling system is designed to dissipate during normal operation
- Example math
  - 10000 H100 GPUs
  - TDP .7 kW
  - Assume 5 requests handled per second.

Data Centers and Water

- Energy Consumption
  - Per hour: 7 MW = 7 MWh
  - Per day: 7 × 24 = 168 MWh/day
- Water
  - 1.5 liters per kWh
  - 168 MWh = 168,000 kWh
  - Water consumption: 168,000 × 1.5 = 252,000 liters/day
  - 66,500 gallons/day
- Are these numbers big? How do we tell?

Data Centers and Water

- Golf Course (Western U.S.)
  - 500,000 – 1,000,000 gallons/day
  - Comparison:
    - Datacenter: ~66,500 gallons/day
    - Golf course: ~8–15× higher
- University Campus
  - 400,000 – 800,000 gallons/day
  - Comparison:
    - Datacenter is smaller but same order of magnitude

Data Centers and Water

- Office Building (Manhattan)
  - 20,000 – 40,000 gallons/day
  - Comparison:
    - Datacenter ≈ 2× larger
- Household (Family of Four)
  - ~300 gallons/day
  - Comparison:
    - Datacenter ≈ 200+ homes

Data Centers and Electricity

- Golf Course
  - < 1 MWh/day
  - Datacenter >> golf course
- University Campus
  - 10s–100s MWh/day
  - Comparable scale
- Office Building
  - Few MWh/day
  - Datacenter ≈ 10–50× larger
- Household
  - ~30 kWh/day
  - 5,600 homes

New AI Models

New Models This Week

- OpenAI Symphony
- DeepSeek-v4
- Talkie

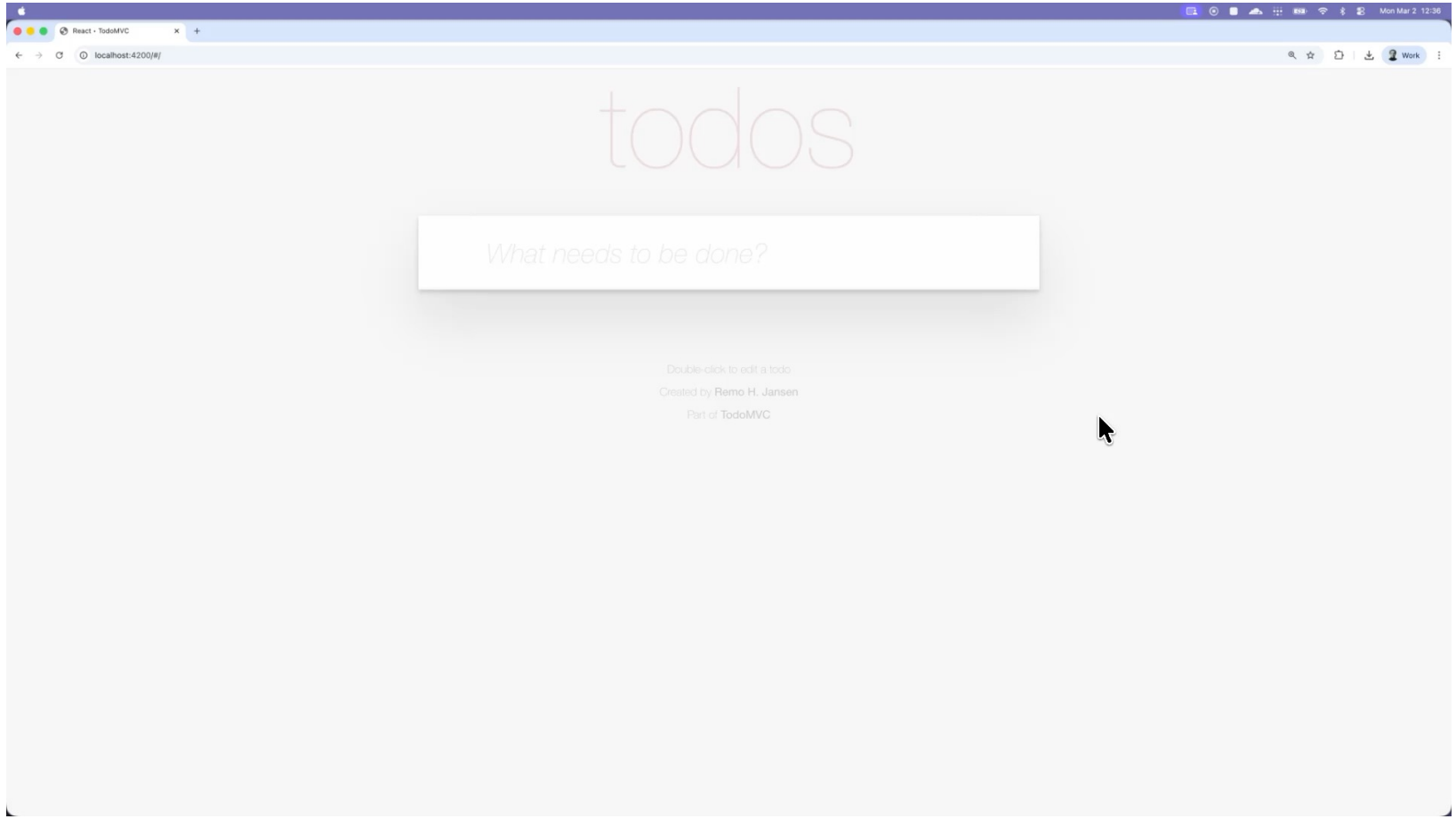
Symphony

- You've used single agents to write code.
- One way we haven't talked about that humans work on code is via *issues*:
  - People create *issues* on a repo to document things that need to be changed to make a project better.
  - https://github.com/trending
- Products like Linear are used for organizing work around issues
  - https://linear.app/

Symphony

- New from OpenAI (April 27)
- Works with Linear: https://linear.app/
- Orchestrates multiple agents to autonomously work on a project.
- Something like this is part of the future of building systems.
- https://openai.com/index/open-source-codex-orchestration-symphony/

Symphony


React • TodoMVC
localhost:4200/#/
Work
todos
What needs to be done?
Double-click to edit a todo
Created by Remo H. Jansen
Part of TodoMVC

DeepSeek-v4

- **DeepSeek-V4-Pro:** 1.6T total / 49B active params.
- **DeepSeek-V4-Flash:** 284B total / 13B active params.
- Both open models
- Both nearly SOTA – Pro Comparable to Opus-4.6-Max and GPT-5.4-xhigh
- If you're going to be around over the summer, I'm setting up a new computer lab with machines that will be able to run at least Flash
  - Let me know if you're interested in helping out/learning
- https://api-docs.deepseek.com/news/news260424

Talkie

- 13-billion-parameter language model
- Trained (almost) exclusively on pre-1931 English text
- Very articulate
- Pretty offensive
- Probably cost < $50,000 to build
- https://talkie-lm.com/chat

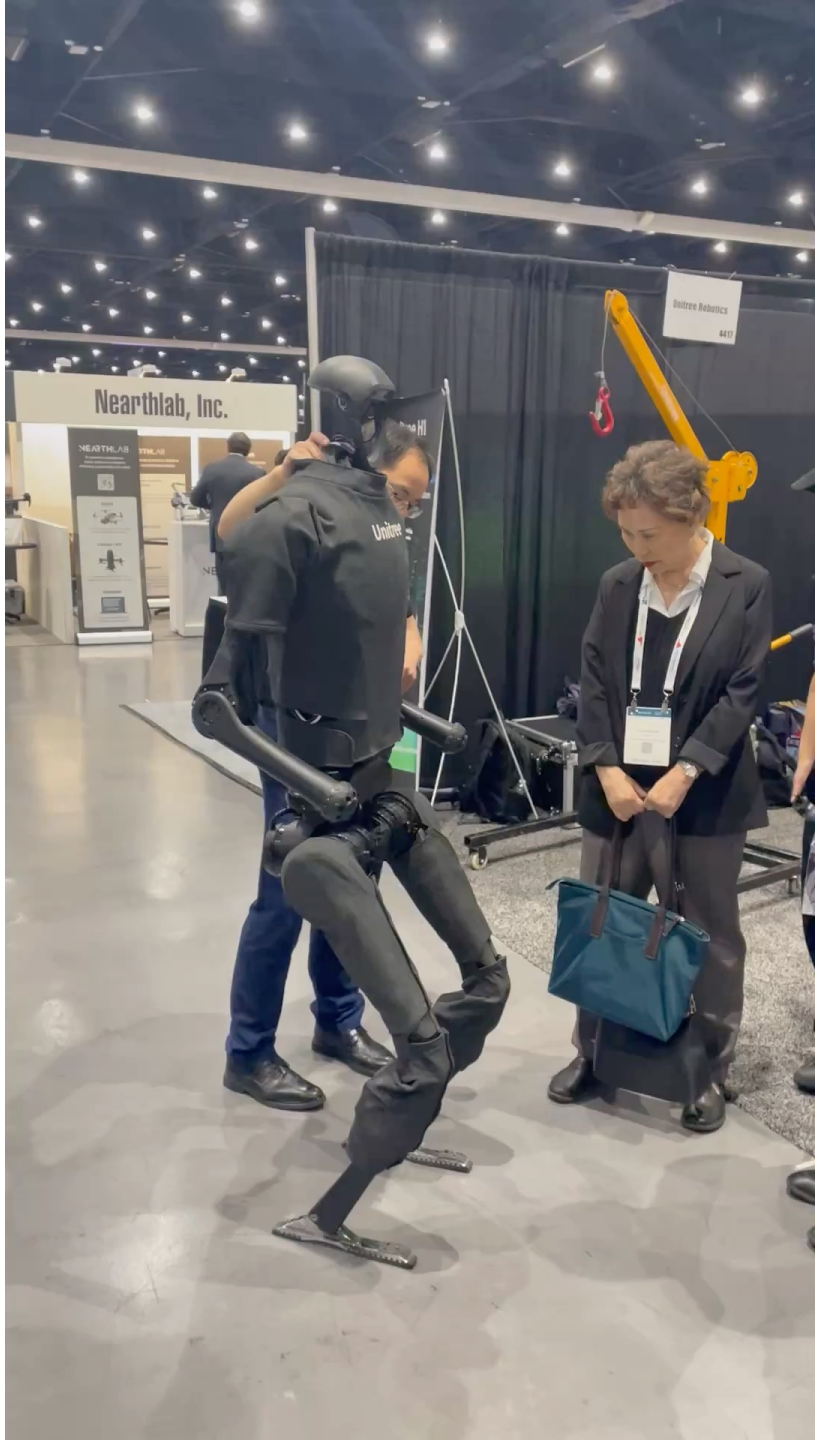
Robotics News

New Robotics This Week

- Humanoids
  - Unitree & OmniXtreme
  - RobotEra
  - Asimov v1
  - JAL “demo”
- Quadrupeds
  - Lynx M20S

Unitree

- Chinese Robotics Company
  - Founded 2016
- Building
  - Robot Dogs
  - Humanoids
- Funding research into autonomous robots
- I recorded this video at AUVSI 2024
  - Major robotics industry event
  - Heavy defense focus

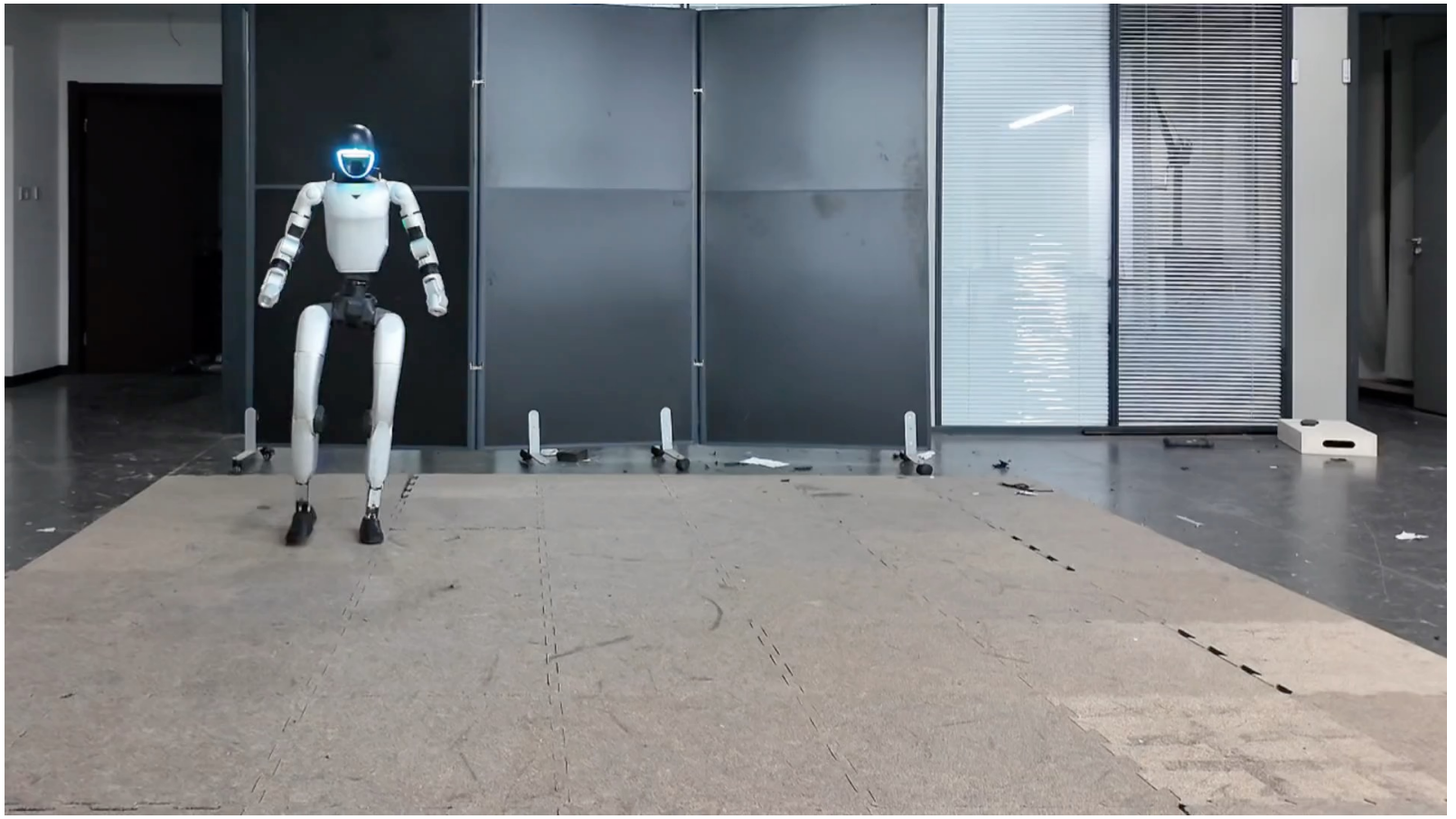


Unitree

- Got their start selling robot dogs.
- Here’s another video I took at the conference
  - I took these videos on April 23, 2024, so almost exact 2 years ago.
  - This was about the limit of their performance at the time.

OmniXtreme

- Beijing Institute for General Artificial Intelligence (BIGAI)
- March 2026
- Just accepted to RSS conference
- https://x.com/RoboPapers/status/2046982108065706239
- https://extreme-humanoid.github.io/



RobotEra

- This is probably where manufacturing and logistics are going.
- Bejing-based company just raised $200M
- https://x.com/CyberRobooo/status/2049118436870259192
- Probably real.

Asimov v1

- Open-source Humanoid
- Only $15,000
- Note: BOM = “Bill of Materials”
  - Manufacturing term
  - List of all the parts you need to build a thing.
- https://x.com/asimovinc/status/2048697696010309842?s=46
- https://manual.asimov.inc/v1

Japan Airlines Haneda Robot “Trial”

- https://x.com/Reuters/status/2049424702536405364
- Watch the video.

Lynx M20S

- https://x.com/DeepRobotics_CN/status/2048704270971859313

- This is probably the optimal form factor for ground-based quadrupeds.
  - You will see lots of this type of robot over the course of your life.

Military Robotics

Reference Points for the Future

- I want to show you some things to give you a sense of where robots (and drones in particular) will likely go.
- The overall trend is towards more autonomy.
  - This is probably bad, and probably inevitable.

Some Chinese Popular Perspective

- These are videos floating around Chinese social media
  - You should assume these are ***all*** propaganda
  - Still useful to see what they're pushing
- They're a bit over the top, but basically accurate
  - Evolution of drone warfare
    - https://x.com/qiaohuanxin/status/2044213342756319401
  - Robots in an invasion of Taiwan
    - https://x.com/GrandpaRoy2/status/2038818052977119457
- These next videos are much more real
  - A couple years ago:
    - https://x.com/alsamahi/status/1855958125729173855?s=20

Chinese News Last Week

Cybersecurity

AI and Cybersecurity

- This is moving fast.
- This is an area where you ***HAVE TO*** protect yourself.
- Good news is that this doesn’t have to be hard.

Mythos

Assessing Claude Mythos Preview's cybersecurity capabilities

April 7, 2026

Nicholas Carlini, Newton Cheng, Keane Lucas, Michael Moore, Milad Nasr, Vinay Prabhushankar, Winnie Xiao

Hakeem Angulu, Evyatar Ben Asher, Jackie Bow, Keir Bradwell, Ben Buchanan, David Forsythe, Daniel Freeman, Alex Gaynor, Xinyang Ge, Logan Graham, Kyla Guru, Hasnain Lakhani, Matt McNiece, Mojtaba Mehrara, Renee Nichol, Adnan Pirzada, Sophia Porter, Andreas Terzis, Kevin Troy

Earlier today we announced Claude Mythos Preview, a new general-purpose language model. This model performs strongly across the board, but it is strikingly capable at computer security tasks. In response, we have launched Project Glasswing, an effort to use Mythos Preview to help secure the world's most critical software, and to prepare the industry for the practices we all will need to adopt to keep ahead of cyberattackers.

https://red.anthropic.com/2026/mythos-preview/

Cybersecurity Terminology

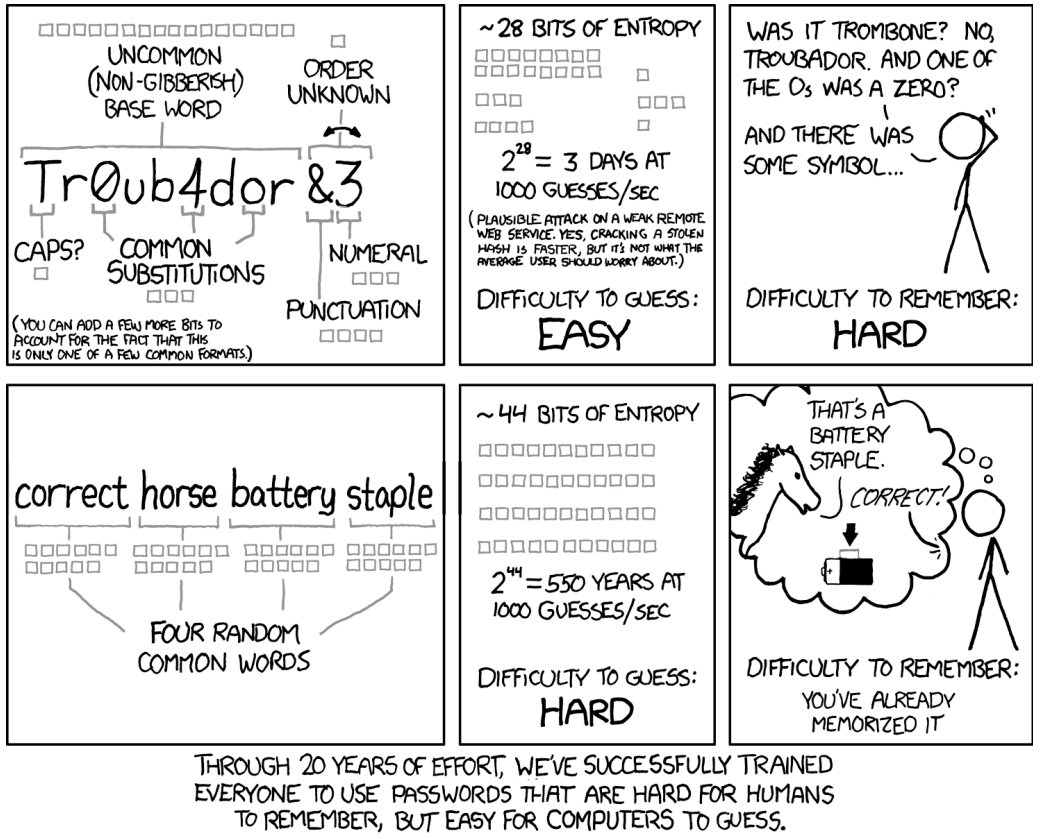
- Zero-day vulnerability
  - A vulnerability or security hole in a computer system unknown to its developers or anyone capable of mitigating it.
- Remote code execution (RCE)
  - The ability to trigger arbitrary code execution over a network.
- Common Vulnerabilities and Exposures (CVE)
  - Reference system for tracking vulnerabilities.
- Mythos shows AI is getting good at finding zero-day and RCE vulnerabilities.
- What can you do? I can point you in some directions.
  - I'll mention software I use, but you should research on your own.

3 Things You Should Know About

- Password managers
- Multi-factor authentication apps
- Hardware authenticators
- These won’t stop all attacks, but they’ll help a lot.

Password Managers

- Program that automatically generates, stores, and autofills passwords for you
- You only have to remember your master password.
- You should be using a password manager.



2FA Apps

- You can’t just rely on good passwords.
- You should be using a two-factor authentication app on every website you care about.
- Google’s Authenticator app is fine. There are other options – investigate them if you don’t want to use Google.

Hardware Authentication

- For things you care about a lot, you should also consider hardware authenticators.
- The most popular option is YubiKey.
- If you use cryptocurrencies, hardware wallets fall into this category.

You need to start thinking about how to secure your digital life

- “Attacks only get better”
- AI will continue to get better at hacking.
- If you start now, you have a good chance of being able to survive most cyberattacks, even from smart AIs.