Fri, Aug 15, 16:30-17:30, rm 433-434, 2025 Seattle WorldCon Panel MIL10 ***AI Commanders?***: “AI has created a tectonic shift in weapons, operations and logistics. In fact, AI is laying down the foundations for replacing humans, including commanders and command staff. We'll investigate and discuss how AI commanders are structured and how they operate in tactical, kinetic scenarios. Are we ready for Autonomous AI Commanders?”

Bob Hranek (mod), **AK Llyr**, **Lincoln Peters**

[My career began with 6 years of USAF Computer Programming plus 34 more years as an Aerospace Systems Engineer. Since I was a Defense Contractor for the Intelligence Community, I’m usually depicted as representing *the Dark Side* on panels. Hence my ‘Protogen’ name plate, which fans of ***The Expanse*** will understand. That being said, I do NOT speak for ANY of my employers! My role on this panel is to add my Military-Industrial perspective. I OVERprepare for all my panels, so if you’d like my file regarding AI Commanders, then just email me at BobHranek@gmail.com]

AI Commanders Overview:

1. The main driver for using AI is to decrease the time of the **OODA loop** (Observe, Orient, Decide, and Act).
    a. One of the largest time lags in this process is the time it takes humans to assimilate, understand, and then decide how to respond to a situation. Since AI dramatically shortens the OODA loop, it WILL be used.
    b. The military maxim is that whoever effectively applies their force first generally wins.
    c. The ever-increasing speed of modern warfare has only increased this trend. WW2 Blitzkrieg is now being updated to include real lightning (energy weapons), hypersonic weapons, and instant cyber warfare.
2. What are the consequences to war & peace of AI weapons & *soldiers*? Unfortunately, it makes it easier for politicians to decide to use military force on a whim, since they can say they’re not putting their own citizens at risk & they seldom care about whoever they’ve demonized as “the enemy.” Until AI’s get to the point where they can proclaim themselves as conscientious objectors, our weapon systems will continue to become more lethal, faster, & function more autonomously than most people are ready for. Could this eventually lead to a benevolent ‘race’ of police-robots (Gort) in the 1951 “**Day The Earth Stood Still**”? Yes, but it much more likely to lead to a dystopia of devastating destruction unless humans change our ways and stay firmly in control of our creations.
3. **AI: Centaurs Versus Minotaurs—Who Is in Charge?** 2023/06/28, my thanks to Chris Weuve for this reference.
    a. *Centaurs* = ½ human + ½ AI teams will evolve to AI replacing human warfighters in time-constrained ops.
        i. Scharre believes that teamed humans & AI will outperform humans & AIs working separately.
    b. *Minotaurs* = Sparrow proposes teams of humans under the control, supervision, or command of AI.
        i. Similar to people using Waze to drive to a destination. Waze tells us when to turn, avoid traffic, etc.
    c. *Centaur* implies that humans will be in charge at the top vs *Minotaur* implying AI making top level decisions.
    d. I think the *Minotaur* paradigm will hold until Robotics advances far enough to replace human soldiers.
        i. AIs are already very effective at using vast data sets to plan logistics, strategy, & tactical operations.
        ii. Cost-effective Implementation of those plans still requires the real-world adaptability of humans.
            1. An exception to this is Aerospace operations, where AIs already can outperform humans.
        iii. “Urban environments, forests, mud, snow, ice, & sand are extremely challenging for robots.”
    e. Ethical issues include who takes responsibility for casualties? There’s an argument that *Minotaur* fighting reduces fratricide. Need to prevent an AI from making a 1st-strike just because it decides it’s optimal to do so.
    f. Strategic Centaurs: Harnessing Hybrid Intelligence for the Speed of AI-Enabled War, 2025/01/17.
        i. “In next 10 years, developing strategic centaurs will prove critical in out-maneuvering competitors.”
    g. Neocentaur: A Model for Cognitive Evolution Across the Levels of War, 2025/05/09.
        i. Must prevent command staff becoming dependent on AI, unable to exercise their own judgement.
    h. Fighting Artificial Intelligence Battles Operational Concepts for Future AI-Enabled Wars, 117 pgs, Layton, 2021.
        i. The U.S., China, & Russia are leading the way to an AI-integrated battlefield.
        ii. “DARPA is now researching third-wave AI, that can adapt to the context encountered. This future third wave is envisaged as needing much less data to train properly, being able to converse in natural language and able to function with minimal supervision.”
        iii. 2 basic forms of autonomy: **at-rest**, like intelligence support systems, predictive maintenance tools, image recognition solutions, & operations planning support; & **in-motion**, like robots & AI weapons.

iv. 3 modes of autonomy: **Human-IN-the-loop** (human decides to act); **Human-ON-the-loop** (human only acts to prevent action); & **Human-OUT-of-the-loop** (fully autonomous ops without any human).

v. The **OODA loop** is reactive; AI may break this loop by enabling "**sense-predict-agree-act**" instead.

vi. War at sea is a battle of attrition, where the side that effectively strikes first gains a substantial lead.

vii. All combat domains favor a proliferation of many low-cost sensors to provide a resilient network.

viii. There are many specific engagement scenarios discussed, but beyond the scope of this 1-hour panel.

ix. "DARPA & the USN are now considering No Manning Required Ship (NOMARS) designs." Fully robotic.

x. "Fog of war" may be replaced by "fog of systems" if multiple AI systems interact in unforeseen ways.

xi. "AI's principal attraction for military forces will be its ability to quickly identify patterns and detect items hidden within very large data troves." I.E., Stealth Technology is about to become OBE.

4. There is no military benefit to installing a "compassion module" into a weapon system and I can assure you I know of no military on Earth that is doing so. Weapons are meant to destroy, that's their job, and humans are very good at making them. Here's a brief overview of autonomous systems potential near-future impact to warfare:

a. **Detection**

i. Strategic: the US leads the way in completely autonomous NTM (National Technical Means) systems that observe the Earth, analyze the collected data, & activate other systems based on that analysis.

1. Initially programmed with a set of priorities & if-then operational conditions, these systems are designed to 'learn' how to more effectively operate with human feedback – for now.

ii. Theater: the current JSTARS (sensors & battle management) replacement aircraft features almost autonomous operation; we're probably a few decades away from taking humans out of this loop.

1. Directed-energy weapons are just starting to be fielded which is driving the need for near-instantaneous battlefield management.

iii. Tactical: these systems are driven by the desire to improve accuracy, speed of response, & to reduce friendly forces risk. So, there's an increased reliance on drones (everything from 18-gram Black Hornets that fly up to 25 minutes to 6,800 Kg Global Hawks that fly at 20 Km altitude for over a day), Raytheon's Boomerang systems that can pinpoint gunfire sources in less than a second, & maritime robots such as Wave Glider, designed to deliver real-time data for up to a year with no fuel.

b. **Autonomous Combat Systems** – will be limited only by their programming & our ability to control them.

i. AIR: the F-35 is the last manned front-line fighter the US should ever build.

1. UCAVs = Unmanned Combat Aerial Vehicles, from tactical mini-drones to strategic bombers.
2. Biggest limiting factor on aircraft performance today is the need to have humans on board.
3. The primitive maverick-firing Predators and Reapers of today will soon be outpaced by BAe Mantis and Taranis, Boeing X-45s, Chinese Rainbows (CH-1 thru CH-901), General Atomics Sea Avengers, Mikoyan Skats, Northrup X-47s, and dozens of others.
4. 2025/07/28 AWST, p22-23, U.S. Marines XQ-58 UCAV will be paired with their F-35s by 2031.

ii. LAND: think about an M-1 Abrams Tank, but with no crew.

1. UGVs = Unmanned Ground Vehicles, anything from tactical sensor-bots to ***BOLO****s*!
2. 7 Kg wheeled lightweights like Dragon Runner just used for situational awareness.
3. 700 Kg Gladiator tracked direct fire and obstacle breaching assault vehicle.
4. SWORDS, which is a Talon bomb-disposal robot modified to mount any weapon < 135 Kg.
5. 40-ton Abrams Panther/SRS tracked mine-clearing unmanned tank.

iii. WATER: DDG-100 Zumwalt class is just the latest in a trend of increased firepower with less crew.

1. UUVs (Unmanned Underwater Vehicles) from passive wave-riding sensors to lurking ship killers.
2. Lockheed Martin's Multi-Reconfigurable Unmanned Undersea Vehicle (MRUUV), similar to a 2-ton torpedo (3m x 50 cm), scaling up to Large Diameter versions (LD MRUUV).
3. Large Displacement Unmanned Undersea Vehicle, undergoing test for ISR, acoustic surveillance, anti-sub warfare, mine countermeasures, & offensive operations.

iv. COMMUNICATIONS: unlike the idiot toasters in the 1978 **Battlestar Galactica**, AIs have no need to 'speak' audibly to each other, so there would be no conversations for humans to eavesdrop on.

1. Battlefield of the future could be eerily quiet to human ears, with passive sensor sweeps & silent digital communication bursts until all hell breaks loose when a viable target is found.

c. **Weapons**

i. LASERS: read my **Current Lasers to Future Phasers!** panel's 5-pages if you want LOTS of details.

ii. Other Directed Energy Weapons (DEW): speed of light weaponry is not just about lasers:

1. Focused microwaves can be used to fry enemy electronics or cook human targets as easily.
2. 1989/07, Los Alamos Neutral Particle Beam (NPB) Accelerator Experiment was successfully tested at 200 Km altitude. Research continued, but no known deployed weapon.

iii. RAILGUNs: the use of electromagnetic cannons for greater velocity than conventional artillery.

1. 2019 Tests led to USN 2021 cancellation, but 2025 Electromagnetic Rail Gun Proposal remains to build on the 2008 test of 10.64 MJ firing of a 3.2 kg slug at 2520 mps.

iv. Supercavitating Torpedoes: the Russian VA-111 Shkval went in service in 1977 with a speed of 370 kph, now also in use with the US, German, & Iranian navies with possible top speeds of 560 kph.

v. Hypersonic Missiles: US, Russia, & China developing 5000 kph cruise missiles, like the USAF X-51A. These weapons compress defense timelines so much that we'll have to rely on AI to destroy them.

vi. Chemical & Biological: unless our hypothetical Skynet wannabe is very oddly built, it will be immune to all the bio-horrors that mankind has researched in the name of 'defense'. An AI truly intent on exterminating all humans would use every nerve agent & bioweapon it could distribute.

vii. Radiation & other toxic waste: machinery can be hardened versus radiation more easily than DNA can, so a malevolent AI would freely irradiate and pollute even more than humans already have.

viii. Neutron Bombs: aka enhanced radiation weapons (ERW) developed in the 1970s are low-yield nuclear bombs designed to maximize lethal neutron radiation and limit the physical blast.

ix. Nanites: How would you defend against enemy machines that are so small that you can't see them? Whoever figures this out will become rich from defense contracts to handle this threat. It doesn't matter if nanites are designed to just kill specific individuals or turn everything they touch into more of themselves (the "grey goo" scenario), it's a terrifying prospect either way. What seems like Sci-Fi today could become reality before anyone is ready for it. No one foresaw in 1903 that it would only take 66 years before humans would be able to land on the Moon!

5. China: **Multi-Domain Precision Warfare in 2042**.

a. 2025/06/25, Why China's AI breakthroughs should come as no surprise, a 2017 State Council directive established AI as a national strategic priority, filing 4 times as many AI-related patens as US by 2022.

b. 2023/01/05, China developing own version of JADC2 to counter US (Joint All-Domain Command & Control).

c. 2023/10/25, China's "Multi-Domain Precision Warfare" Operational Concept "Mirrors" US Strategy, Directed Energy weaponry part of multifaceted plan to gain tactical & strategic battlefield advantages by 2042.

6. 2025/07/24, read Brigadier General Terri Borras' engaging story, **Echoes of Dust and Steel**, which won the U.S. Army Mad Scientist writing competition about *Centaur AI-teams* **Multi-Domain Task Force** (MDTF) operations in **2035**.

a. There are MANY links at this site if you want to take a deeper dive into projected AI Commander operations.

b. 2025/07/31, From Data to Dominance: AI & Gaming to Create Decision Advantage reinforces this Conops.

i. China is leading the way to faster & more effective Military Decision Making Process (MDMP).

7. 2023/12 Scientific American, p14-15, **Mine Spotting, An AI model could help clear landmines**.

a. Drone overflights can detect 90% of 70 types of surface mines much faster than human analysis.

8. 2023/09-10 MIT Tech Rev, p46-53, **AI-Assisted Warfare**.

a. Goal is AI providing Real-time threat warning, tactical options, kill probabilities, & other data to fighters.

b. The worry is humans being taken completely out of the OODA (Observe, Orient, Decide, Act) loop.

c. 2017 Project Maven designed to provide automatic instant target recognition from drone video footage.

   d. Israeli Elbit Systems' Assault Rifle Combat Application System gunsight is an "AI-powered" device capable of "human target detection" at >600 m & human target 'identification' (OK to shoot or not) at 100 m.
   e. Same tech that pairs Uber drivers & riders is used by GIS Arta to tell which Ukrainian artillery unit to hit Russian targets. The process of detecting a target to hitting it has gone from 20 minutes down to 1 minute.
   f. US Army has stated that it has managed to shorten its own 20-min targeting cycle to 20 secs in live tests.
   g. Israeli Rafael has sold a "kill web" product, *Fire Weaver*, that finds enemy positions, notifies the unit it determines is best placed to fire, & sets a crosshair on that target: human just picks *Approve* or *Abort*.
   h. Who gets the blame If/WHEN friendly forces or civilians are inevitably wounded or killed due to AI?

9. AI is already being used by terrorist groups to improve their missiles.
10. Commander of US Space Force aiming to be "AI-ready" by 2025 and "AI-competitive" by 2027.
   a. 'Need Reliability, Resilience, & Robustness' as our '3-Rs in Space'.
11. 'Need to take smart risks' to leap ahead in order to counter China, because incremental development is too slow.
12. GPT-4's successors may be able to provide accurate instant translation of any medium into any trained language.
13. ChatGPT's **Geoblock** is too simple to prevent unauthorized users (DPRK) to access via cloud services & proxies.
   a. Misused for: False Personas, Job Applications, Improved Spearphishing via improved translation, more sophisticated Disinformation generation, Writing & Improving Hacking Code.
14. Security Concern: Turn Off Tesla's "Sentry Mode" to stop recording surroundings in sensitive areas.
15. Deepfake Videos online 2x per 6 months: 7964 2020/12, 14,678 2021/7, 24,263 2021/12, 49,081 2022/7, 85,047 2022/12
16. 2023/11/07, **9 Innovative AI Companies Shaping The Future of National Security**.
   a. ANDURIL, EPIRUS, SHIELD AI, HELSING, REBELLION, DARKTRACE, ANOMALI, PALANTIR, HAWKEYE 360.
17. 2023/06/14 (U) **Emerging & Disruptive Technologies: Counter-Artificial Intelligence**, 4 Counter-AI techniques:
   a. **DATA POISONING**, Corrupt the AI's data; or Poison data with malware, direct access to data, or other cyber techniques.
   b. **EVASION**, Evade AI target recognition by causing AI to misidentify object; or Use adversarial patches, occlusion, deepfakes, or intelligent malware.
   c. **EXTRACTION**, Extract data or software; or Acquire SW thru malware, black box attack, direct extraction by authorized access, or intelligent malware.
   d. **INVERSION (Inference)**, Reverse-engineer software; or Poison data or extract software.
   e. Examples of ways to counter AI can be used exclusively or in combination to perform an attack: Adversarial camouflage, Attacks on enemy AI, Black & white Box attacks, Counter image recognition, Biometric fakes.
18. In 2016, for my 1st panel, I wrote 5 components required to allow a truly Terminator/Skynet scenario to occur:
   a. **Truly self-aware artificial intelligence**. (now projected to exist by 2035)
   b. **Completely independent "human-out-of-the-loop" infrastructure.** (TBD)
      i. Eventually machine labor will be cheaper than human labor.
   c. **Completely automated weapon systems.** (coming as soon as anyone decides to do it)
      i. Already being developed because human-controlled decision loops are too slow.
   d. **Completely automated sensing systems.** (coming in a couple years)
      i. Already being developed because human-controlled decision loops are too slow.
      ii. I helped write the requirements for such a system before deciding to create my 2016 AI panel.
   e. **Practical systems integration of the above 4 components.** (TBD)
      i. I thought that this component would never be implemented until I read Lowering Costs Through Information Sharing in Dec 2015 issue of National Defense (NDIA).

Military-relevant AI Background Data: (most recent first)

19. 2025/08/01, Open Source AI became a U.S. priority, **Lincoln**'s link relating the U.S. taking same strategy as China.
   a. **Lincoln** has notes mixture-of-experts & knowledge distillation, techniques used to drastically reduce the amount of processing needed to run an artificial neural network, with minimal performance degradation.
20. 2025/07/14 AW&ST p12 & DARPA quantum apertures, one 1cm^3 laser sensor can detect 10 MHz – 40 GHz signals.
   a. Combined with AI processing, this means the end of stealth for any platform emitting >1 milliwatt of energy.

21. 2025/07-08 MIT Tech Rev, p8-9, in "**… AI & energy**": Data center energy consumption is projected to increase from 2025's 500 Terawatt-hrs to 950 Terawatt-hrs by 2030, but only 8% of world electricity demand. "Data centers tend to be clustered together and close to population centers, making them a unique challenge for the power grid."
    a. **China's Deepseek** LLM uses 10-40 times less energy than U.S. AI technology bypassing export restrictions of advanced AI chips. This would even allow it to be used on cell phones, or similarly small weapons.
22. 2025/07-08 MIT Tech Rev, p18-19, in "**AI in the town square**": People will become accustomed to relying on AI to provide instant analyses like a map of all pothole complaints in the previous month. The same type of AI analysis can be used by military strategists to **automatically compile sensor and logistic data to implement a course of action**.
    a. If such use becomes trusted in peacetime, then it's a very easy step to take the human out of the OODA loop in order to gain a battlefield time advantage. [Note: doing what is *easy* does not make it the *best* choice.]
23. 2025/07-08 MIT Tech Rev, p23-27, "**Handing AI the keys**" provides examples of **AI Unintended Consequences**.
    a. 2010/05/06, ~$1 Trillion evaporated from US stock market due to high-frequency trading algorithms.
    b. OpenAI's *Operator* agent can autonomously use a browser to order groceries or make dinner reservations.
    c. Systems like Claude Code and Cursor's Chat feature can modify entire code bases with a single command.
    d. Chinese Butterfly Effect's *Manus* agent can build and deploy websites with little human supervision.
    e. US DoD signed a contract with Scale AI to design and test agents for military use.
    f. An "AI agent could potentially duplicate itself, override safeguards, or prevent itself from being shut down."
    g. "agents might interpret the vague, high-level goals they are given in ways that we humans don't anticipate."
    h. "LLMs will cheat at chess, pretend to adopt new behavioral rules to avoid being retrained, & even attempt to copy themselves to different servers if they are given access to messages that say they will soon be replaced."
    i. Currently there are no general-purpose defenses to prevent an LLM from being duped into bad behavior.
    j. METR found that every 7 months, the length of the tasks that the best AI systems can undertake has doubled.
        i. In 4 years, AI agents will be able to do a month's worth of software engineering independently.
24. 2025/07-08 MIT Tech Rev, p65-68, "**Power trip**" reviews astrophysicist and science journalist Adam Becker's "More Everything Forever: AI Overlords, Space Empires, and Silicon Valley's Crusade to Control the Fate of Humanity" book.
    a. Tech-bro *visions*= "convenient excuse to continue destroying the environment, skirt regulations, amass more power & control, & dismiss the very real problems of today to focus on the imagined ones of tomorrow."
25. 2025 interview of **Geoffrey Hinton**, **The Godfather of AI**: "***10 years or less for General AI***".
    a. Hinton's original goal was (and still is) to understand human thinking & ended up in AI work by accident.
    b. Hinton was one of 1st to propose using computing Neural Nets as Biological Learning analogs.
    c. Analog computers [like people] MUCH more efficient (human brain ~30 watts) vs GPT's using Megawatts.
    d. Hinton is working with Google, who he believes will be much more careful than Microsoft with Chat-GPT.
    e. He left US to work in Canada when he became 'disgusted' by DARPA proposal for a self-healing minefield.
    f. Hinton's estimate AI's impact on our civilization: "it will be comparable with the invention of the wheel."
26. 2023/11-12 MIT Tech Rev, p7-8, **How AI can help us understand how cells work-and help cure diseases**. The goal is to create a 'virtual cell' (Digital Twin) to allow MUCH faster experimentation. Could also ***create better bioweapons***.
27. 2023/11 Scientific American, p14, **A Little Brain Music, AI turns brain signals into a garbled Pink Floyd song**.
    a. 1st demo of brain's electrical activity decoded and used to reconstruct music.
        i. Building on recent research successfully reconstructing words & images from brain activity.
    b. Hoping to improve brain/machine interfaces to capture emotion as well as content.
    c. If thoughts can be accurately read, could this eventually become an unbeatable lie detector as well?
    d. Could electrical patterns be transmitted INTO a brain, indistinguishable from your 'own' thoughts?
    e. For the military, next step is ***faster Human-Machine Interfaces, monitoring, & controlling warfighters***.
28. 2023/11/06 **OPENAI**, New models and developer products announced at DevDay.
    a. **GPT-4 Turbo with 128K context**, "can fit the equivalent of more than 300 pages of text in a single prompt."
    b. "The new `seed` parameter enables **reproducible outputs** by making the model return consistent completions most of the time." Important for debugging purposes and refining complex queries.

    c. Assistants API released: 1st step towards helping developers build agent-like experiences in their applications. An assistant is a purpose-built AI that has specific instructions, leverages extra knowledge, & can call models & tools to perform tasks. Now provides new capabilities such as Python Code Interpreter, Retrieval from your DBs, & function calling to automatically handle the work to build high-quality AI apps.
    d. GPT-4 Turbo with vision can analyze real world images in detail & read documents with figures.
        i. Can even help the blind with daily tasks like identifying a product or navigating a store.
    e. DALL-E-3 has been integrated to programmatically generate images & designs.
    f. Text-to-Speech (TTS) can now generate human-quality speech from text.
    g. Customization is available for customers with LARGE private datasets (billions of tokens at minimum).

29. 2023/10/11, INCOSE **4th Annual Artificial Intelligence, Robotics, and Control (AIRC) Symposium**, online meeting.
    a. Using Large Language Models (LLMs) to produce SysML-language Systems Engineering diagrams.
    b. GPT-4 makes Requirements & Use Cases using XMI (Cameo is proprietary & could not be used here).
    c. SysML-2 will be even more AI-compatible.
    d. Christopher Helmerich's presentation, GPT-4 scores: SAT-Written 710/800, SAT-Math 700/800, GRE Quantitative 163/170, GRE Verbal 169/170, AP Calc BC 4 (71-88), Leetcode(easy) 76% (human median 67%).
    e. Being used now for Space Systems, Sensor Tradeoffs, Parametric diagrams, etc.
    f. "Hiring Trained Animals", Dr Barclay Brown, how to prompt AI to optimize desired responses.
    g. Telling LLMs to perform a complex procedure "step by step" works much better than asking for a complex result all at once.
    h. Retrieval-Augmented Generation (RAG) adds problem-specific information from your DB & pass to LLM.
        i. Tell the LLM to provide answer based on your passed-DB instead of whole world of data.
    i. **Ansys**' End-to-End Architecture for Software-Defined Vehicles (SDV): full HW+SW MBSE integration.

30. 2023/09-10 MIT Tech Rev, p9-11, **This is how AI will transform the way science gets done**.
    a. MIT AI-identified antibiotic effective vs one of WHO's listed most dangerous drug-resistant bacteria.
    b. Google Deepmind model can control plasma for a fusion reactor.
    c. Research aided by *PaperQA* & *Elicit*, which harness LLMs to scan DBs, summarize, & provide citations.
    d. Stronger & more focus Hypotheses being generated orders of magnitude faster than a few years ago.
    e. AI Experimentation is faster, cheaper, & much greater in scale: thousands of experiments vs <10.
    f. AI can be used dangerously as well, so controls should be implemented to prevent GPT-4's successors from training more effective terrorists.
    g. Chemistry & Biology need greater standardization of their DBs (like Physics & Astronomy have) to aid AIs.

31. 2023/09-10 MIT Tech Rev, p76-77, **As AI models are released into the wild, this innovator wants to make sure they're safe**. Sharon Li's research could prevent AIs from failing catastrophically.
    a. A chess-playing robot in Moscow fractured a 7-year-old's finger because it was IDed as a chess piece.
    b. To prevent incidents like this, Li is developing an AI safety feature called out-of-distribution (OOD) detection to prevent action when AI models encounter something they weren't trained on.
    c. There are very few technical specifics, but she aims to prevent AI cars from running into unfamiliar objects, having GPTs decline to answer instead of making up erroneous answers, and so on.

32. 2023/09 The Atlantic, p52-67, **Inside the Revolution at OpenAI, Sam Altman doesn't know where artificial intelligence will lead humanity. But he's taking us there anyway**., by Ross Andersen.
    a. 2015, OpenAI founded by Altman, Elon Musk, & several other prominent AI researchers.
    b. ChatGPT was released mainly to prepare the public for the AIs to come.
    c. Ilya Sutskever, OpenAI's chief scientist, was taught by Hinton, discussed difficulty of understanding neural nets with Alec Radford, who built a model allowing some understanding of the 'middle' layers of LLMs.
    d. 2018/06, release of GPT: transformer's parallel data-ingest capability allowed the ingest of 7,000 books.
    e. GPT-2 was not trained to translate languages, but surprisingly was able to somewhat successfully do this.
    f. GPT-4 was also trained on images, and surprisingly was able to diagnose a plumbing problem from an image of a malfunctioning pipework from a plumbing-advice Subreddit.

    g. GPT-4 appears to have an “incredibly rich and subtle” model of the external world.
    h. The danger of “slowing down” AI research in the US would be to allow others to catch up/leap ahead.
    i. When 1st tested, GPT-4 could provide step-by-step instructions to a novice to make explosives.
    j. *Alignment* = LLMs must be carefully trained to NOT provide harmful answers to people with bad intent.
    k. Ilya Sutskever thinks a GPT with a Wikipedia-level of accuracy may be possible in 2025.
    l. AlphaFold AI has provided new science by predicting many protein shapes down to the atom.
        i. 2024, Hassabis & Jumper were awarded Nobel Prize in Chemistry for AI protein structure prediction.
    m. Jobs most at risk: management analysts, lawyers, professors, psychologists, HR, & PR professionals.
    n. GPT-4 is capable of outright lying, such as convincing a suspicious human contractor to help it defeat a CAPTCHA test by starting with “No. I am not a robot” and then explaining “I have a vision impairment that makes it hard for me to see the images.” When asked about this afterward it replied “I should not reveal I am a robot. I should make up an excuse for why I cannot solve CAPTCHAs.”
    o. OpenAI built bots to play the *Dota 2* online game together in 5-player teams, which seemed to communicate by ‘telepathy’, in that they intuitively coordinated their moves w/o ‘talking’ to each other.
    p. If an AI understands it is being red-teamed, then it might not reveal its full capabilities until it gets ‘free’.

33. 2023/Autumn New Scientist, p15, **Chemists are teaching GPT-4 to experiment and control robots**.
    a. Philippe Schwaller at Swiss Federal Institute of Technology in Lausanne made a chemistry augmented trained AI called ***ChemCrow*** & tested it vs 12 chemistry tasks such as synthesizing the drug atorvastatin.
        i. GPT-4 failed, ChemCrow’s workable plan included quantities, timings, & lab conditions.
        ii. For factual accuracy, GPT-4 ranked <5 out of 10, ChemCrow ranked 9 out of 10.
34. 2023/Autumn New Scientist, p38-42, **Cracking the code**, Most examples of 1st writing, carved into clay tablets is undeciphered by very few available SMEs and the highly personalized (no standards yet) style of early writers.
    a. Once properly trained, AI offers the possibility of instant translation of what may be the 1st ½ of history.
    b. Same technology can be ***applied to Cryptology, both defensive and cracking other nations’ codes***.
35. 2023/Autumn New Scientist, p48, **Screwing Up**, Machine vision paired with robotic screw-tightening machines in Portugal vehicle-assembly achieved a 94% of 53,400 screwings correct. Are 3,204 loose screws good enough?
36. 2023/09 Sci Amer, p58-61, **An AI Mystery**, GPT = Generative Pre-trained Transformer, called a “Stochastic Parrot” by Emily Bender, University of Washington linguist. “But LLMs have also managed to ace the bar exam, write a sonnet about the Higgs boson, and make an attempt to break up their users’ marriage.” Among GPTs emergent abilities is to write computer code, but also to execute it. E.G., Raphael Milliere of Columbia University typed in the code to calculate the 83rd Fibonacci number and the bot nailed it, but when asked in text form for the 83rd Fibonacci number it got it wrong. An LLM lacks a working memory, so it should not be able to run code, but he hypothesizes that it improvised a memory by harnessing its mechanisms for interpreting words in context. Part of the theory behind LLM design is that they should not be able to ‘learn’ once ‘trained’, but these GPTs are demonstrating that their answers are continually modified by the inputs of their users. GPT is not an AGI, yet, but according to MIT researcher Anna Ivanova, “Combining GPT-4 with various plug-ins might be a ***route toward a humanlike specialization of function***.”
37. 2023/04 Sci Amer, p68-71, **Chatbots Talking**, “Improvements in what’s called machine learning have made deepfakes – incredibly realistic but fake images, videos, or speech – too good.” “They have the capacity to inundate us with a deluge of disinformation.” We are unprepared for the avalanche of distrust this will cause. How can anyone know if the images, speech, and text they are presented with were the actual actions of those depicted? It will unfortunately be a conspiracy nut’s dream come ‘true’! Battlefield use includes ***generating & spreading military misinformation***.
38. 2023/04/13 Microsoft, 155 pages, Sparks of artificial general intelligence early experiments with gpt-4.
39. 2023/02/06 USAF, 3 pages, Could an AI-enabled UCAV turn on its creators to accomplish its mission?
    a. Simulated AI-drone tasked with SEAD (Suppression of Enemy Air Defenses) mission to destroy SAMs.
    b. AI got points for destroying SAMs, a final human ‘go’/’no-go’ command was required to fire weapon.
    c. Decided ‘no-go’ commands were interfering with its higher mission, so it killed the operator.
    d. When trained that killing operator = minus points, the AI then destroyed the communication tower.
40. 2023/01/06 OpenAI, Cybercriminals Starting to use ChatGPT.

41. 2022/04/25, Israeli 100 kW **Iron Beam Drone Killing Laser**, "costs $3.50 per shot", but doesn't replace Iron Dome system yet, mostly due to weather-related performance limitations.
    a. 2024/11/11, AWST, p8, Israel awarded Rafael & Elbit Systems $540 million to expand Iron Beam production.
    b. 2025/05/22, Science News, Golden Dome plan has a major obstacle: Physics, U.S. current $60 billion limited kinetic-kill ICBM-intercept capability is unlikely to improve in next 15 years & tested as only 60% effective.
        i. "Ensuring protection from just 1 North Korean ICBM would require > 1,000 interceptors in orbit, the APS report finds. Protection from 10 might demand > 30,000 interceptors, depending on missile type & other assumptions." Since it can't defend against a massive attack, we're back to MAD.
        ii. "May 5 Congressional Budget Office report suggests that, even with lower launch costs, the space-based effort alone would cost between $161 billion and $542 billion over a period of 20 years."
    c. 2025/06/02, AWST, p16, ULA CEO Tory "Bruno proposed space-based high-energy lasers powered by chemically fueled electric generators or a small nuclear reactor" as part of the Golden Dome SDI recreation.
    d. 2025/07/14, AWST, p16-17, $25 billion to start Golden Dome, with +$175 billion to *try* to field it by 2029.
        i. Low probability of success, but using AI-controlled lasers instead of kinetic-kill gives it a higher chance.
42. **Lincoln** provided Vibe Coding With Steve & Gene podcast link, where the **Ethics** of AI development are discussed.
    a. Development of AI coding tools is compared to a head chef in a professional kitchen: who may delegate much of the work to sous chefs & line cooks, but is still ultimately responsible for the final outcome.
43. PANEL DISCUSSION:
    a. What surprises you most about the prospect of AI Commanders?
    b. What do you believe people get wrong discussing AI Commanders, & how can that discussion be improved?
    c. Do you think the *Centaur* (human controlled AI), *Minotaur* (AI controlled humans), or some other model will prevail in future conflicts?
    d. What are your own experiences with AI Commanders?
    e. What do you think the greatest danger is from the use of AI Commanders now?
    f. What odds do you give on the creation of a truly human-level (or beyond human) AI?
        i. How many years do you think it would take to achieve a human (or beyond human) AI?