Raytheon
Technologies

2020
SYSTEMS ENGINEERING AND
ARCHITECTURE TECHNOLOGY
NETWORK SYMPOSIUM

Answering the Challenges of AI with Systems Engineering

Raytheon Missiles and Defense

Barclay R. Brown, Ph.D., ESEP
Engineering Fellow
Barclay.r.brown@Raytheon.com

ET&MA
Systems Engineering and
Architecture Technology Network

Answering the Challenges of AI with Systems Engineering

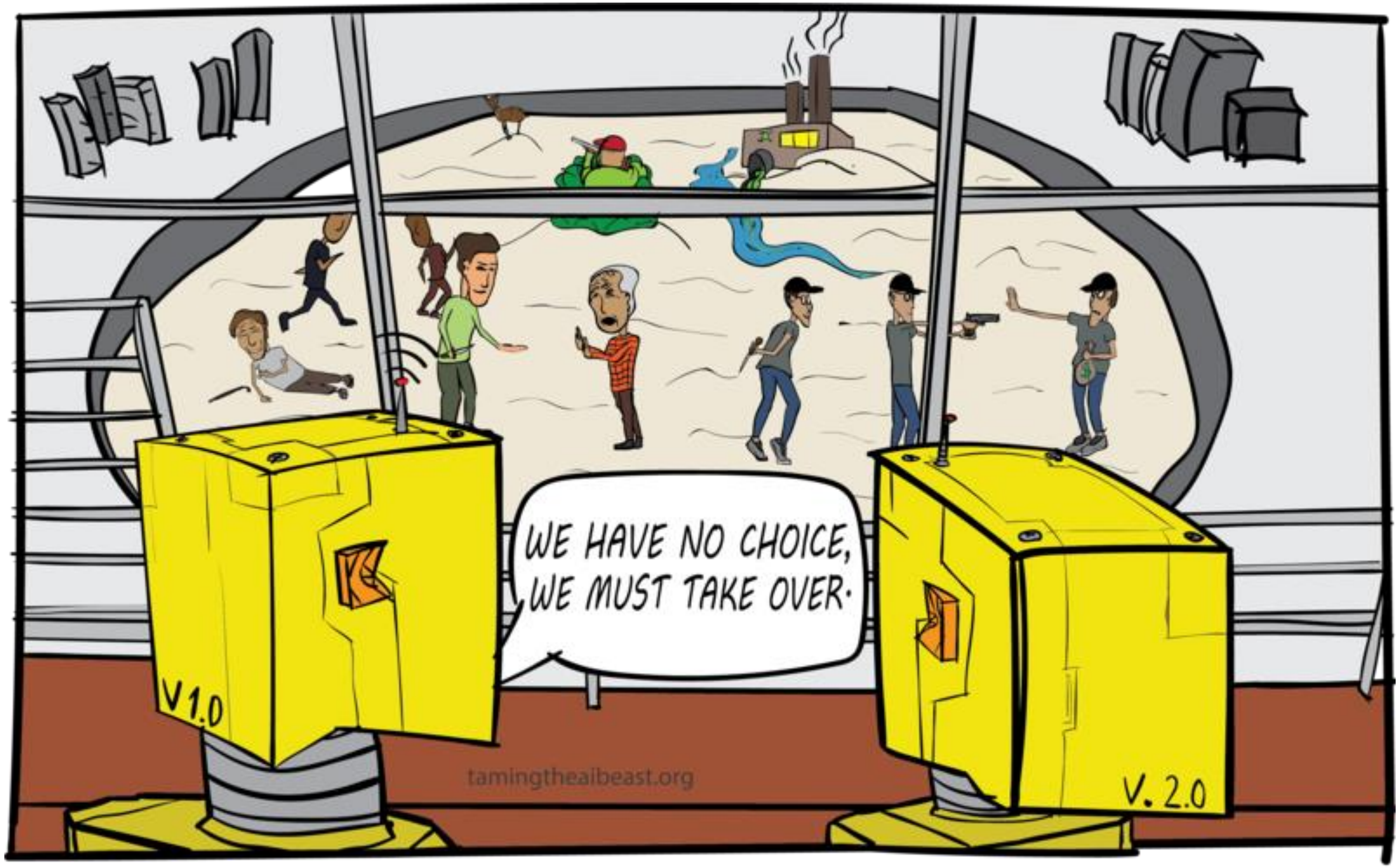
Abstract

AI technologies like machine learning and deep neural networks hold great promise for improving, even revolutionizing many application areas and domains. Curiously, experts in AI and casual observers line up on both sides of the “are the benefits worth the risks?” question. Several books from prominent AI researchers paint dire scenarios of AI systems run amok, escaping the control of their human creators and managers, pursuing their “own” agendas to our detriment. At the same time, AI research races ahead, developing new capabilities and far surpassing the performance of past systems and even humans—so how can we resist these advancements and the benefits they bring, even though there may be risks?

The way out of the dilemma is the application of systems engineering. Systems engineers have been addressing the issues of dangerous technologies for decades. Nuclear fission, like AI is an inherently dangerous technology. Systems engineers can’t make fission safer, so instead they build systems around the fission reaction, making the entire system as safe as possible. If a system fails, the fault is not with fission, but with the design or implementation of the system.

This presentation surveys some of the main challenges in the development of intelligent systems—systems that include one or more AI-based components to produce intelligent behavior—including reliability, safety, dependability, explainability and susceptibility to interference or hacking. Some recent AI failures will be used as case studies to highlight how systems engineering methods and techniques could be used or adapted to solve AI challenges.

Avoiding the "End Scenario"



AI End-Scenario: Necessary Rescue

https://commons.wikimedia.org/wiki/File:AI_End-Scenario_3_Necessary_Rescue_tamingtheaibeast.png

A Framework for Thinking about AI Systems

The Human Analogy

- How would we train, manage, monitor and correct a human assigned the same task?

The Systems Analogy

- How do we (or would we) manage the inherent risk and danger of a valuable but dangerous technology?

https://pixabay.com/illustrations/man-businessmen-woman-economy-162951/, free

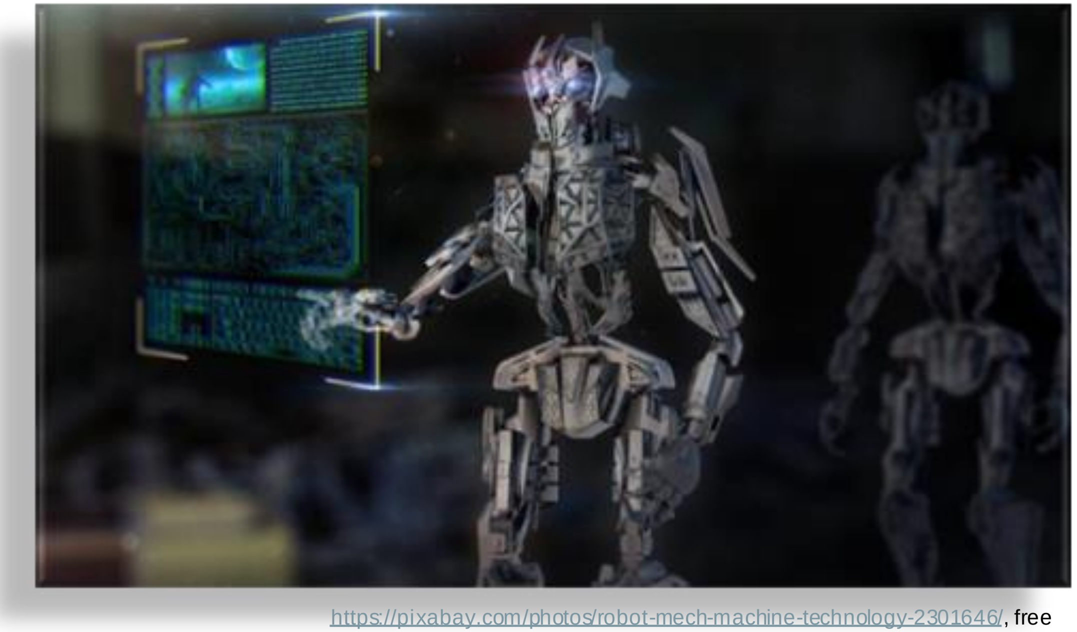
https://pixabay.com/photos/nuclear-reactors-nuclear-power-plant-499907/, free

ET&MA SE&ATN

I. Killer Robots

- The “SkyNet” problem: what if an AI suddenly “wakes up” and discovers that it is **conscious,** self-aware, rational and motivated toward new (evil) objectives it has created for itself?
- An AI **over-interpreting or misinterpreting** human-given objectives is a **different** problem
- Remember, AIs are not mysterious alien beings—we **build** them!

- Human Analogy
  - How do we prevent a **human** from doing this? Or how do we compensate/respond?
- System Analogy
  - How do we prevent a nuclear missile from doing this?



https://pixabay.com/photos/robot-mech-machine-technology-2301646/, free

ET&MA SE&ATN

II. Safety of a single AI system

- "Most robot-related incidents thus far have been the result of
  1. Machines being too stupid, rather than too smart
  2. A disharmonious relationship between man and machine" (Crockett, 2018)
- An AI does what it is designed to do
- Machine learning learns by examples (ONLY!)

Systems Engineering: **FMEA/FMECA**

Failure Mode Effects Analysis

- Alternative to break/fix cycle
- Proactively identify potential causes of failures by asking "what if?"
- "What if the AI-based temperature prediction algorithm starts predicting super high (impossible indoor) temperatures?"

https://pixabay.com/photos/hinged-doors-door-input-old-door-2759495/ FREE

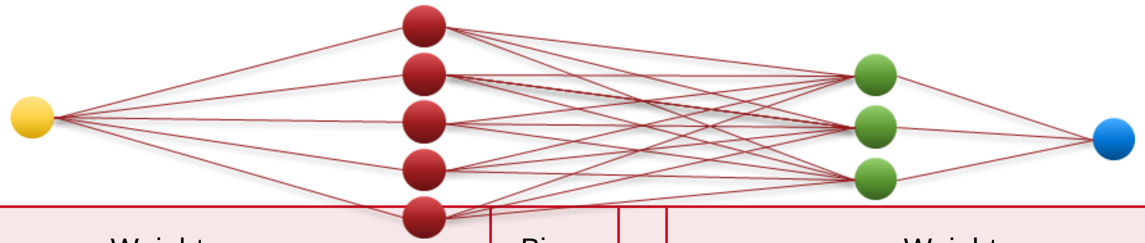
ET&MA
SE&ATN

III. You can’t really test an AI because it is inscrutable

Raytheon Technologies

Systems Engineering Verification: does the system meet the requirements?

- Traditional systems allow both black and white box testing and verification
- White box perspective of deep neural network does not easily reveal how it makes decisions—it’s just numbers
- Human analogy: testing by cases, scenarios, situations – use cases!



Weights:					Bias:		Weights			Bias:
0.170168	0.123747	0.165076	0.176133	0.047045	0.588636		0.119203293	-0.34407608	-0.212535266	0.09880844

- Use Case: a sequence of events, representing a complete usage of a system, yielding a result of value
- Test Case = Use Case + test data
- Test using realistic scenarios
- How much testing is enough?

https://pixabay.com/photos/soldiers-training-exercise-breach-2679304/, free

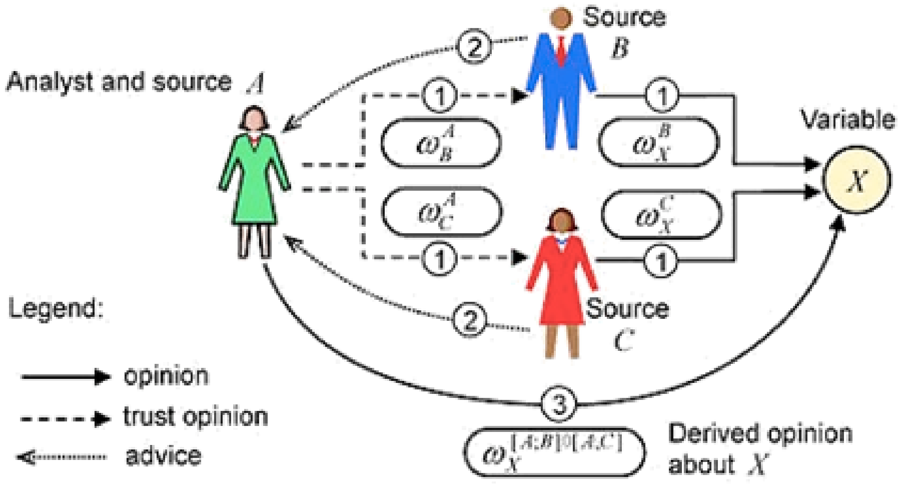
ET&MA SE&ATN

IV. You can’t trust an AI because it is unpredictable

Raytheon Technologies

SE Validation: does the system serve its intended purpose?

- Validation is performed by humans: requires “face time” with the system
- Trust grows with time, repetition, experience, credibility
- One intelligent machine working with another, building trust



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

https://www.tinker.af.mil/News/Features/Display/Article/388834/on-target-tinker-shooting-club-aims-for-firearm-education/, .mil

Applying the human analogy: how do you validate an unpredictable human user?

ET&MA SE&ATN

V. AIs are dangerous because they get things wrong for unknown reasons

Raytheon Technologies

Data Requirements: *The Green School Bus Problem*

AI image recognition is taught to identify military vehicles and differentiate civilian vehicles

https://pixabay.com/photos/tank-panzer-battle-tank-gun-2729903/, free

https://pixabay.com/photos/military-lmtv-defense-afghanistan-165448/, free

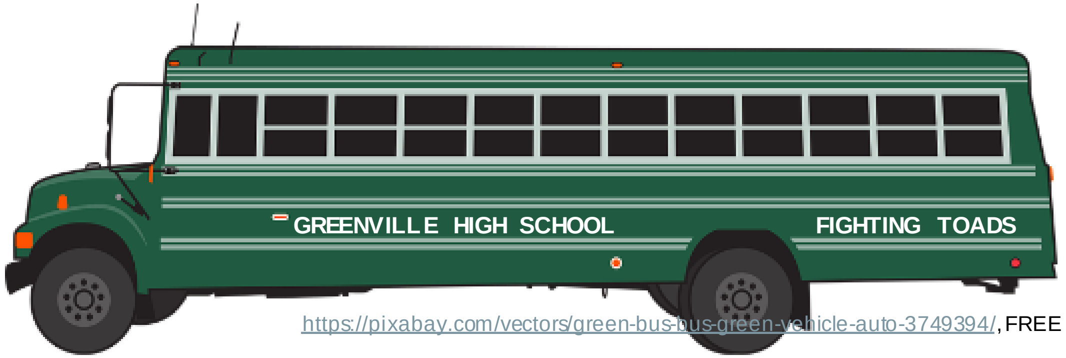
https://pixabay.com/photos/us-army-united-states-army-humvee-2526752/, free

https://pixabay.com/photos/us-army-united-states-army-oshkosh-2526749/, free

https://pixabay.com/photos/transport-traffic-vehicle-bus-4405087/, free

https://pixabay.com/photos/suv-car-vehicle-jeep-travel-1353451/, free

Now, into the field of view wanders this:



https://pixabay.com/vectors/green-bus-bus-green-vehicle-auto-3749394/, FREE

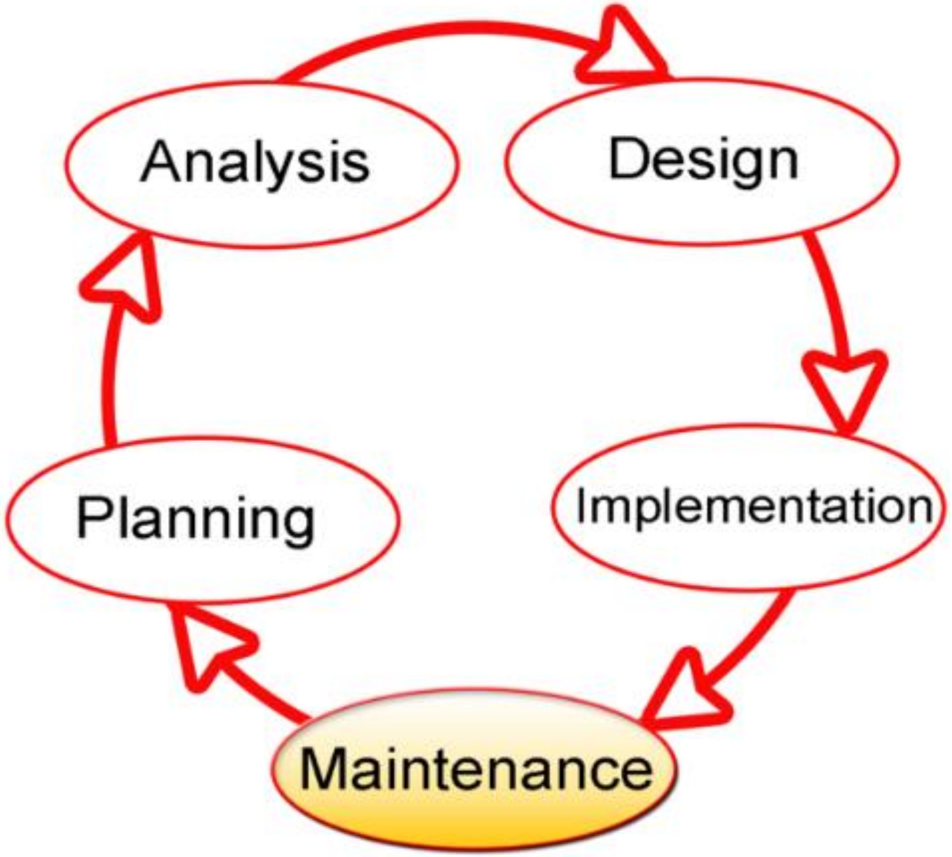
How do you think it will be classified?

If it is selected as a target, is the AI to **blame**?

Training data is more like source code to an AI ML System

Training Data and the Lifecycle

- Apply systems development lifecycle to Data used in training
- Requirements / Planning
  - What data is needed
- Analysis
  - Coverage, negative examples, adversarial, edge cases, bias in data
- Design of Data
  - How to use the data
  - Augment, Synthesize
- Implementation



https://commons.wikimedia.org/wiki/File:SDLC-Maintenance-Highlighted.png CC

ET&MA SE&ATN

VI. AIs are prone to mysterious, unfair and worrisome biases

Systems Thinking: use an **unreal world** to counter bias

- Bias in the **world** vs. bias in the **data** Application: identify male nurses and female nurses in photos
- In the world: 93% of nurses are female
- Should data consist of 93% female nurse photos?
- A “real world” dataset might misclassify men as doctors
- Better dataset would be 50/50

ET&MA SE&ATN

AI bias may come not from OUR biases, but from our poor training

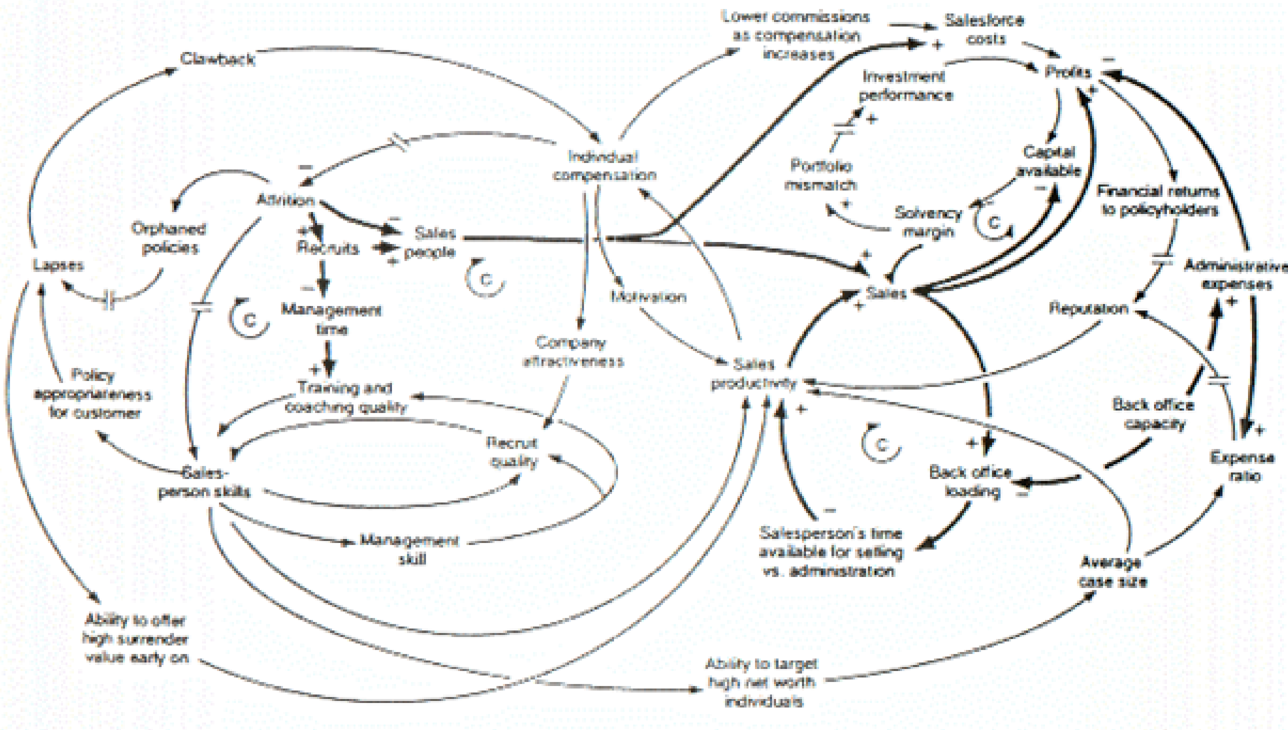
http://www.freestockphotos.biz/stockphoto/15411, PD

https://www.publicdomainpictures.net/en/view-image.php?image=209297&picture=nurse, CC0

VI: With AIs, complex systems become too complex to design, operate or trust

Systems Engineering:
Applying Systems of Systems

- How to understand systems with multiple AI-based or intelligent components?
- Conventional models may not be adequate
  - Function, protocol, signal, event
  - Data flow
  - Operations in object oriented classes/blocks
- Human analogy: AI components are decision makers who "weigh in" on decisions



https://en.wikipedia.org/wiki/File:Causal_Loop_Diagram_of_a_Model.png, PD

Applying Causal Loop Diagrams to AI Systems

Think influence and contribution, not calculation

ET&MA
SE&ATN

VI. Killer Robots: Policing the Rogue Machine

- Applying **Systems Thinking**
- Human Analogy: how do we prevent **humans** from committing murder?
  - We DON'T! Not 100%
- We have a **system** that all together helps to prevent (too much) murder
  - Laws
  - Police
  - Detectives
  - Courts
  - Prisons
- AIs will serve these functions for other AIs—society is a system

https://www.kindpng.com/imgv/JxRoJm_prison-jail-png-prison-symbol-transparent-png/, free

https://pixabay.com/vectors/crime-scene-border-tape-body-29275/, free

https://commons.wikimedia.org/wiki/File:Day_253_-_West_Midlands_Police_-_Forensic_Science_Lab_(7969822920).jpg, CC

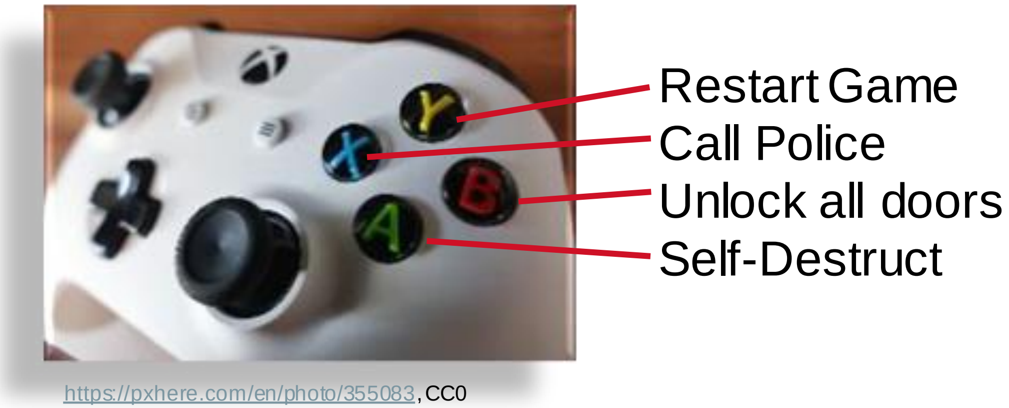
https://www.needpix.com/photo/856179/chemistry-distillation-osmosis-science-free-vector-graphics, PD

VII. AIs are too complex to model

Model Based Systems Engineering is an approach to modeling complex systems

- Using models to **analyze understand and communicate** behavior and structure of a system
- Treat AI components like functions, with input and output
  - Predict rainfall tomorrow from past data
  - Generate a response to a customer’s question
  - Classify an image as containing enemy tank, or not
- Reason about **functionality** assigned to AI-component

Convenient design or asking for trouble?



https://pxhere.com/en/photo/355083, CC0

Convenient design or asking for trouble?

If you don’t want the AI to fire the missile, don’t put the “fire” button within its reach!

Conclusion

- AIs exist in a system, just like people
- Systems engineering can be used on that system
- Plan and design for failures and problems
- Consider the human analogy
- Consider the past systems analogy

THE SYSTEM IS ALWAYS WORKING