Sandia National Laboratories

Exceptional service in the national interest

Human/AI Collaboration
Dancing with the Bear

Paul Schutte (pschutt@sandia.gov)

APPLIED COGNITIVE SCIENCE (5572)

Presented to the NASA Engineering Safety Center

August 8th, 2023

Key takeaways …

- Behind every successful AI, stands a strong human

- A design that counts on humans to have machine strengths
    or machines to have human strengths is poor design

- Human Machine Teaming is about the right allocation of functions and tasks between humans and machines
    – not just an easy interface and usability

- Automation should comply with Crew Resource Management Principles

- Metaphors can be useful in design, training, and usability

- Sometimes the best design will give tasks to the human that the machine might do better,
    in order to improve the human's situation awareness.

Outline

- Will Humans become Obsolete?

- AI, the Dancing Bear – Shall we Dance?

- Handicapping the Human by Design

- Training the Bear to be a Team Player

- Metaphors in human/machine teaming design

- Putting it all together

- A final question for you

Will Humans become obsolete?

"The next wave of economic dislocations won't come from overseas. It will come from the relentless pace of automation that makes a lot of good, middle-class jobs obsolete."

Former President Barack Obama

"AI labs [are] locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control."

Pause Giant AI Experiments: An Open Letter with over 33000 signatures

"Humans will always be required on the flight deck. Someone has to be there to cycle the circuit breakers."

Anonymous Airline Pilot

"I honestly think you ought to sit down calmly, take a stress pill, and think things over."

HAL 9000.

Let's Ask Google Bard

- What human jobs will you replace?
  - Customer service representatives
  - Data entry clerks
  - Writers

- What human jobs will AI never be able to replace?
  - Creativity
  - Empathy
  - Social Skills
  - Critical thinking

- *It is possible that AI will someday be capable of doing all of the things listed above, but it is impossible to say for sure. AI is a rapidly evolving field, and it is difficult to predict what the future holds.*
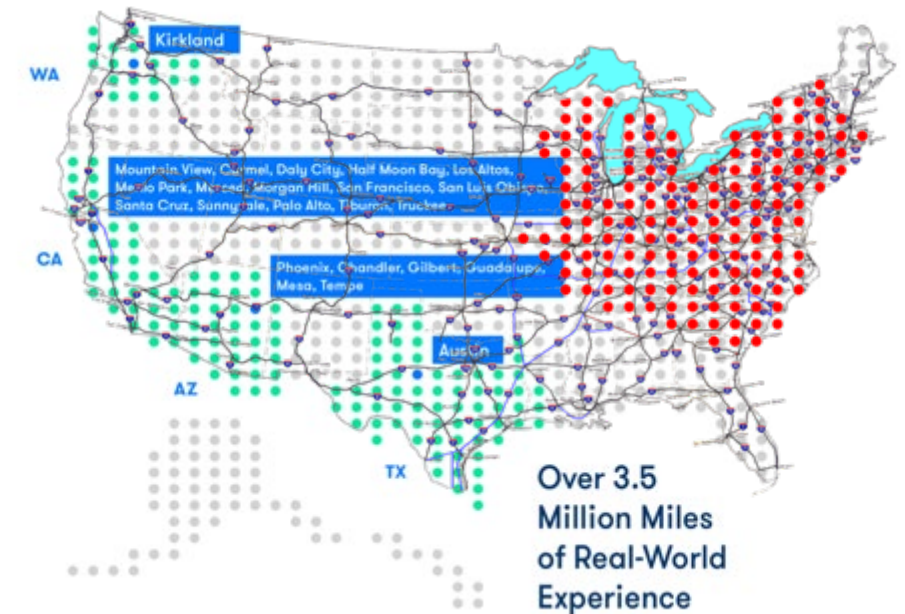
Foundation Models, Deep Neural Nets, and AlphaGo, Baby!

- AI is doing amazing things!
  - AlphaGo developed a new strategy that no human had thought of!
  - A large language model wrote a entirely new scene of M*A*S*H that was performed by Alan Alda and Mike Farrell
  - AI can create graphics from text descriptions
    - (most of the graphics in this talk were created using less than 10 word prompts!)
  - Cars can drive autonomously
  - AI can accurately summarize papers and news articles
  - AI can provide relatively simple natural language explanations of complex phenomena
  - There are hundreds of books available on Amazon that were written entirely by a large language models

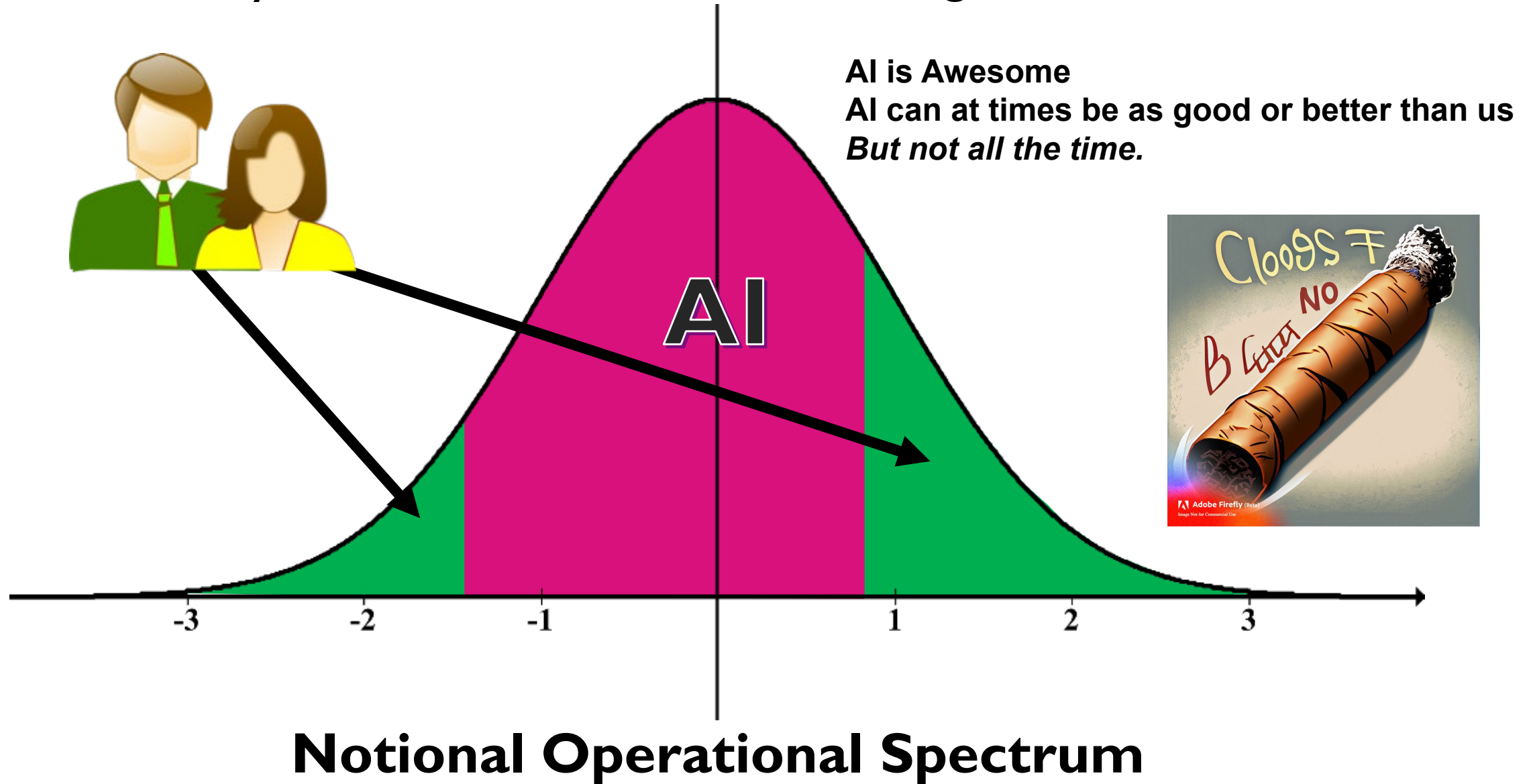- It won't be long now before AI can do anything a human can do, right?

Foundation Models, Deep Neural Nets, and AlphaGo

- It won't be long now before AI can do anything a human can do, right?
  - Humans play games like Chess and Go because they are hard for us, not because they represent what humans do best. Games have rules you can't break. Life? Not so much.
  - Alan Alda said the AI "has a terrible sense of humor" and came up with jokes like "My grandmother once bet on a horse that turned out to be a cow and still managed to make a profit."
  - Often, AI generated art is goofy when you want serious...
  - Yes, AI can drive cars, but where have the cars been driven?
  - Large Language Models are very good liars, pro
    It is common enough to have its own label, "Ha
  - There are no AI books on any best seller lists.

Over 3.5 Million Miles of Real-World Experience

Behind every successful AI, stands a strong human

AI is Awesome
AI can at times be as good or better than us
But not all the time.

AI

Notional Operational Spectrum

AI - The Dancing Bear

"the wonder isn't that the bear
dances **well**
but that the bear dances at all"
Alan Cooper
The Inmates are Running the Asylum

AI is absolutely amazing
because it's not a human

But what if it had a human
dance partner?

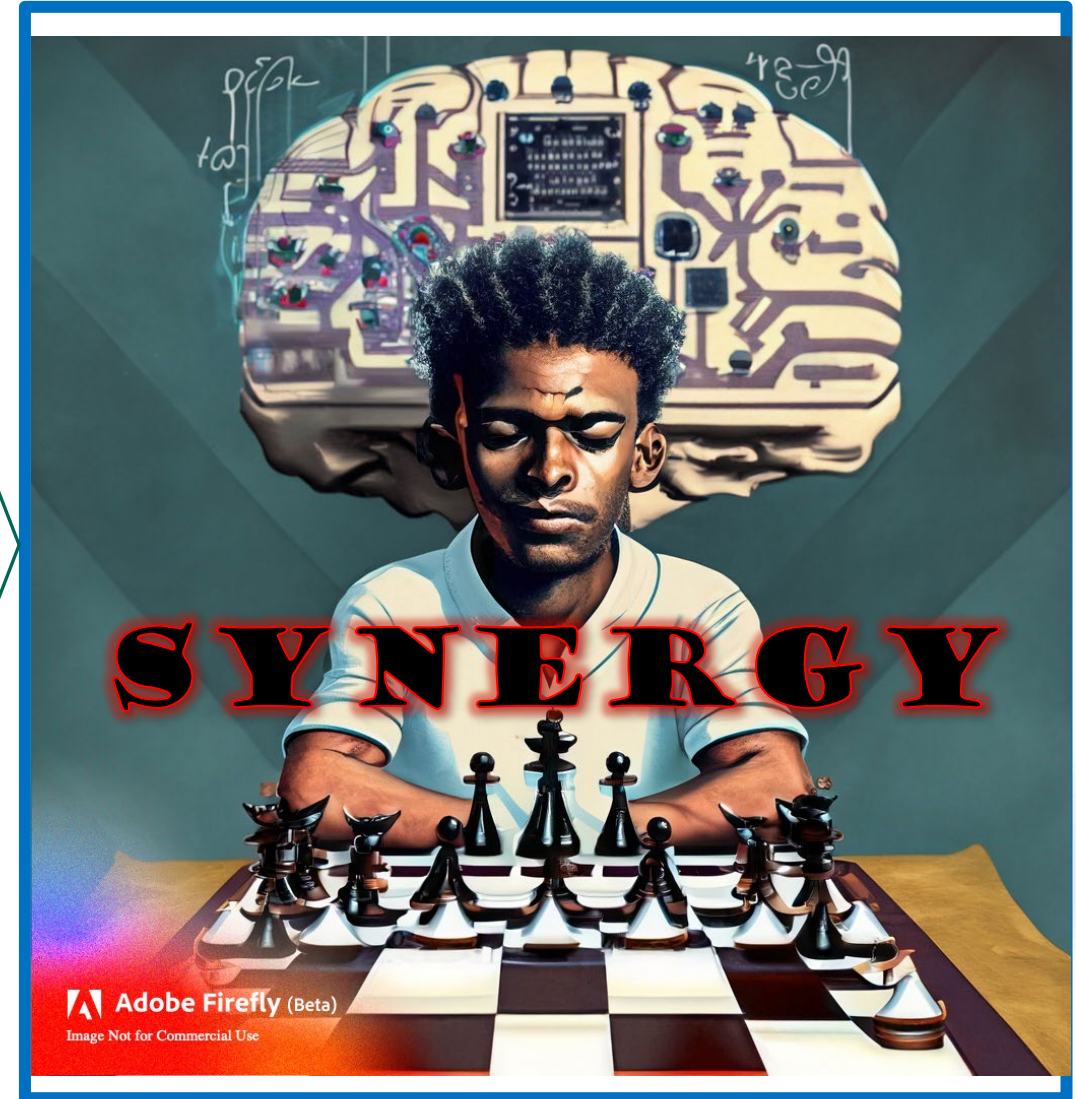Adobe Firefly (Beta)
Image Not for Commercial Use

Shall We Dance?

Human/Machine Teaming - *Better together*
Example – 2005 Freestyle Chess Tournament

GM Ability
2500-3000

AI

Were All
Defeated
By
These
Guys

SYNERGY

Adobe Firefly (Beta)
Image Not for Commercial Use

Some Do's and Don'ts when Dancing with Bears

Humans' Great Moves

Engage the human to keep them situationally aware

- Creative
- Adaptive
- Goal Driven
- Multipurpose
- Multimodal
- Teachable
- Efficient
- Empathetic
- Intuition
- Will-to-Live

Bears' Great Moves

- Memory
  - Declarative
  - Retrospective
  - Prospective
  - Procedural
- Computation
- Vigilance
- Endurance
- Not Distractible
- Standardized
- Speed
- No Human Errors

Do not step on the bear's toes

Adobe Firefly (Beta)
Image Not for Commercial Use

A design that counts on humans to have machine strengths or machines to have human strengths is _poor design._
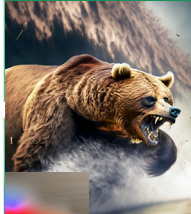
12

When Bears Get Mean – Handicapping the Human by Design

When Bears Get Mean - Example

SAE INTERNATIONAL

SAE J3016™ LEVELS OF DRIVING AUTOMATION™

Learn more here: sae.org/standards/content/j3016_202104

Copyright © 2021 SAE International. The summary table may be freely copied and distributed AS-IS provided that SAE International is acknowledged as the source of the content.

	SAE LEVEL 0™	SAE LEVEL 1™	SAE LEVEL 2™	SAE LEVEL 3™	SAE LEVEL 4™	SAE LEVEL 5™
What does the human in the driver's seat have to do?				When the feature requests, you must drive		
	These are driver support features				hated driving features	
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering OR brake/ acceleration support to the driver	These features provide steering AND brake/ acceleration support to the driver	the vehicle ... and will not operate unless all required conditions are met		This feature can drive the vehicle under all conditions
Example Features	• automatic emergency braking • blind spot warning • lane departure warning	• lane centering OR • adaptive cruise control	• lane centering AND • adaptive cruise control at the same time	• traffic jam chauffeur	• local driverless taxi • pedals/ steering wheel may or may not be installed	• same as level 4, but feature can drive everywhere in all conditions

14

Why? Let's ask Science! Modified Yerkes-Dodson Law

Normal Operations – 90% of the time – the bear drives

The human must monitor the <u>bear</u> and the <u>situation</u> (e.g., the road)

Requires humans to be vigilant, not distracted but not engaged.

Maximal

Minimal

Maximal

Minimal

When the feature requests, you must drive

KEEP AN EYE ON THE BEAR

TAKE OVER WHEN THE BEAR CAN'T

USUALLY HIGH STRESS SITUATION

Hypostress

Stress Level

Hyperstress

Hancock & Warm 1989

OMG! It's almost like it's designed to handicap the human!

A design that counts on humans to have machine strengths or machines to have human strengths is *poor design*
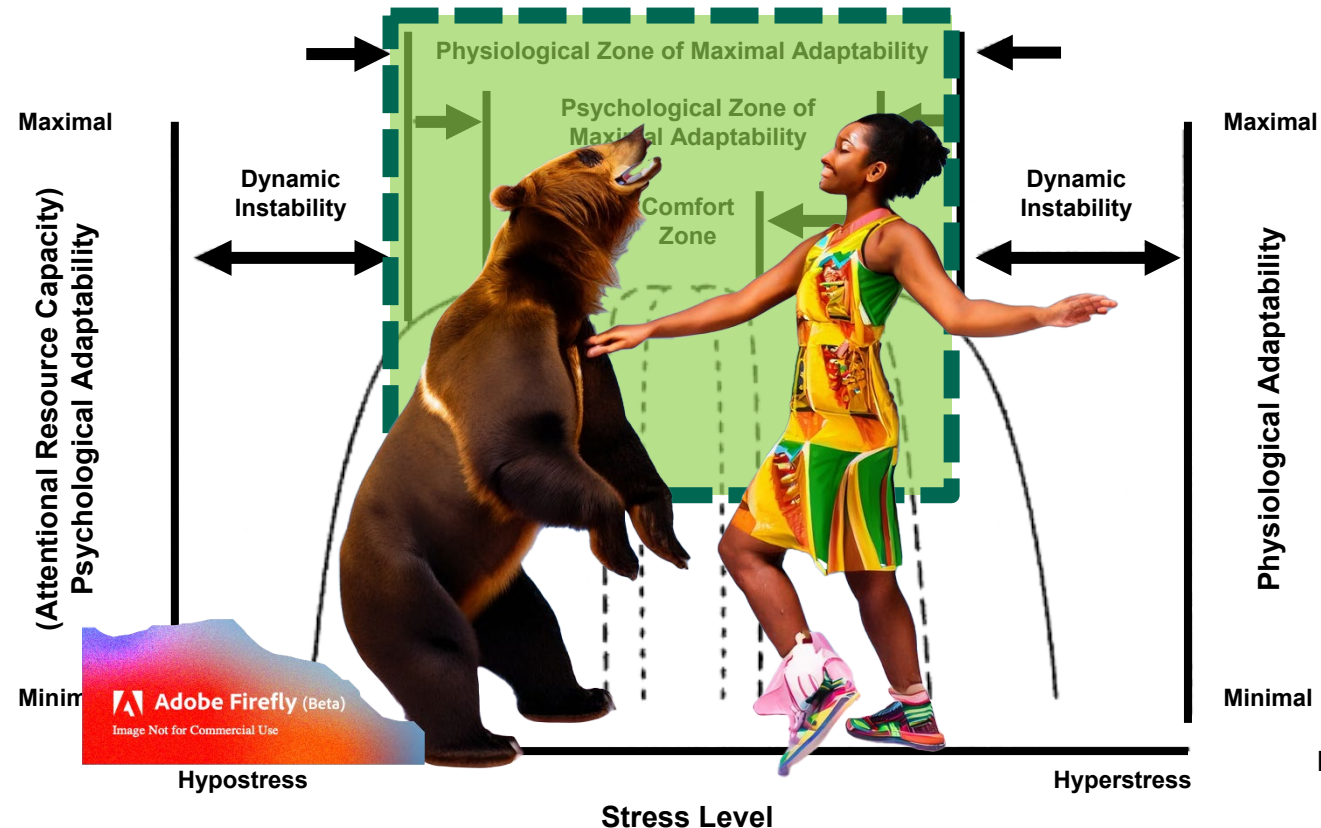

Hancock & Warm 1989

Dancing in Flow – Keep the Human in the Dance!

This may mean allocating tasks to the humans that the automation could do (better)

Human Machine Teaming is about the right allocation of functions and tasks between humans and machines – not just an easy interface and usability.

Hancock & Warm 1989

Training the Bear to be a Team Player

- Crew Resource Management
  - Developed in response to multiple accidents where perfectly good aircraft crashed
  - Investigations showed that lack of coordination, communication, willingness to speak up, etc. among the crew was a major causal factor
  - Developed a training package to teach crews how to work as a team in normal and non-normal situations
  - It has been EXTREMELY EFFECTIVE!!

- Traditionally, Automation is a poor CRM participant!
  - Strong, Stubborn, Silent
  - If a human pilot behaved like automation, they'd be fired.

Assertiveness

Situation Awareness

Flexibility

Mission Analysis

Communication

Decision Making

Leadership

Automation and CRM

- Hold Automation to a higher standard!
  - Automation should comply with relevant aspects of CRM[1]
  - But as a machine teammate, not as a human teammate

- For example:

- Communicate when something is abnormal and be specific

- Positive Control – automation can request that the human take over but it should wait for positive response from human before automation relinquishes control

- Task assignments – Humans and automation should be able to share the task load and switch tasks gracefully when necessary

- Monitor teammates. Humans still need to monitor the automation but computers should constantly monitor humans for signs of fatigue and ill-health.

[1]Shively, R. J., Lachter, J., Koteskey, R., & Brandt, S. L. (2018). Crew resource management for automated teammates (CRM-A). In Engineering Psychology and Cognitive Ergonomics: 15th International Conference, EPCE 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings 15 (pp. 215-229). Springer International Publishing.

Metaphors
Automation can be a Bear, … or a Horse, or a Sheepdog

- Automation can be a good team member and adhere to CRM *without using natural language*
  - Many high-functioning teams try to limit their vocabulary
  - Humans can confuse the meaning of natural language

- Metaphors can be useful in design, training, and usability
  - *The car behaves like a well-trained horse*
  - Design – set a template, how would you do this on a horse? How would the horse behave?
    - The designer doesn't have to do it that way, but it's a conscious design decision to vary from the template
  - Training – If you can tap into a knowledge schema that the user already knows, then you just need to correct and/or modify that schema to fit the design. They already know most of it
  - Usability – Framing the automation performance in a familiar or natural context may make predicting how the automation will behave more intuitive, even if it hasn't been trained

- Let's talk more about metaphors, but first …

Let's say goodbye to the Bear Metaphor

Metaphors[1], there not just for Desktops anymore

- **Target** IS LIKE A **Source** – Life is like a Box of Chocolates

- But you need more. How is the Target like the Source? – You never know what you're gonna get

- Types of metaphor 'Hows'
  - Functional – The Target is performing the function of the Source
    - Trains are like Iron Horses. ATMs are like robot Tellers. Microwaves are like super-fast gas ovens
  - Communication – The Target communicates with the user like the Source
    - Siri, Alexa, Google communicate like a human. The alert barks at you like a dog
  - Behavioral – The Target behaves like the Source
    - The Pilot's Associate behaves like a Human Copilot. This car behaves like a horse. This robotic arm behaves like your arm
  - Internal – The Target works just like the Source
    - The heart works like a pump. The Neural Network works just like the neurons in your brain.

[1]For all you semanticists out there, Analogy might be more appropriate than Metaphors, however since the term 'desktop metaphor' is widely used in computer science and human factors, I'm sticking with Metaphors

Different Sources for Metaphors

- Human, Domesticated Animal, Body Part, Plant, and Tool

- It's important to know that each of these Sources can be said to have a form of intelligence
  - It can be different from human intelligence
  - Can help set user expectations regarding intelligence and capability

- Warning – communication metaphors can imply behavioral metaphors
  - Natural Language may imply human intelligence

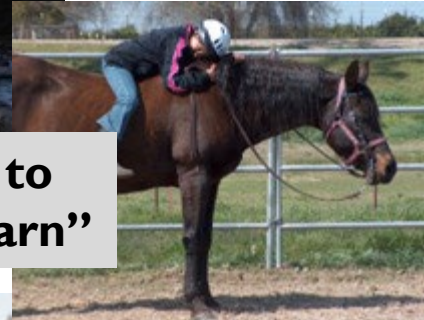|	Human	Animal	Body Part	Plant	Tool
Functional	Bank Teller => Automated Teller	Horse & Wagon => Train	Arm => Wrench	CO2 absorption => Carbon Filter	Typewriter => Word Processor
Communication	Human speech => Voice Synthesis	Bark => Klaxon	Nervous System => Neural Controller	Observation	Typewriter => QWERTY keyboard
Behavioral	Copilot => Pilot's Associate	Horse => *SAFEdeck*	Human Arm => Robotic Arm	Photosynthesis => Solar Cells	Telephone => Internet Phone
Internal	Nervous System => Neural Networks	Dog Legs => Robot Legs	Hypothalamus => Thermostat	Branching => Database Structure	Pipes => UNIX Data Transfer
Group	Platoon	Herd	Fingers	Garden	Notebook

Example - Horse Metaphor[1] for aviation or automobiles

Environment Adaptation
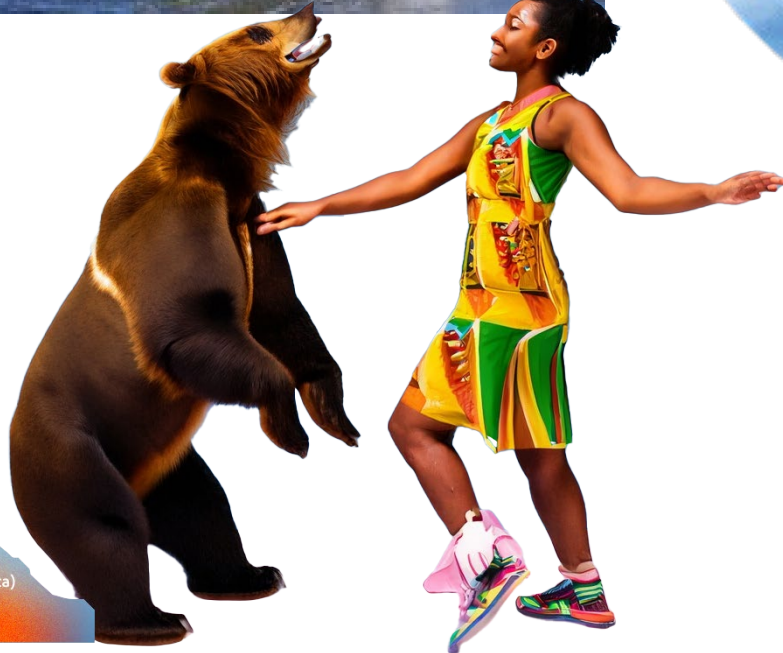
Path Following

"Back to The Barn"

Self Preservation

Self Preservation Overruled

Temporary Autonomy

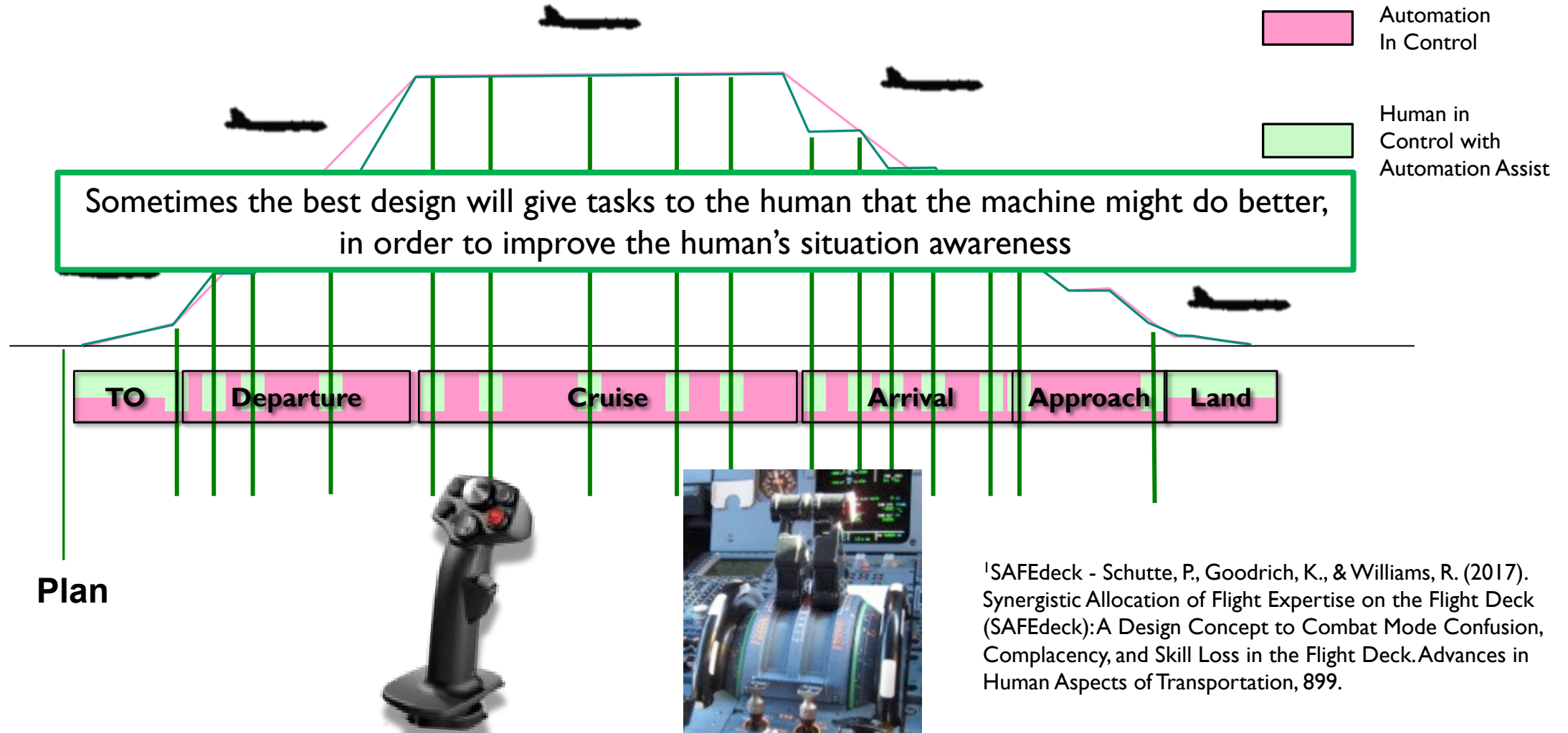[1]Flemisch, F. O., Adams, C. A., Conway, S. R., Goodrich, K. H., Palmer, M. T., & Schutte, P. C. (2003). The H-Metaphor as a guideline for vehicle automation and interaction (No. L-18448).

24

Putting it all together

SAFEdeck[1] Example

Pilot Makes All Major Course Changes Through Inceptors
Automation Performs Behaviors

Automation In Control

Human in Control with Automation Assist

Sometimes the best design will give tasks to the human that the machine might do better, in order to improve the human's situation awareness

| TO | Departure | Cruise | Arrival | Approach | Land |

Plan

[1]SAFEdeck - Schutte, P., Goodrich, K., & Williams, R. (2017). Synergistic Allocation of Flight Expertise on the Flight Deck (SAFEdeck): A Design Concept to Combat Mode Confusion, Complacency, and Skill Loss in the Flight Deck. Advances in Human Aspects of Transportation, 899.

SAFEdeck Example

PILOT

- *Responsible* for Flight
- *Performs* all gross maneuvers
  - Maintain Situation Awareness
  - Retain skills
- *Monitors*
  - Flight Progress
  - Overall System Health
  - Automation Behavior
- *Delegates* Tasks
  - Flight Crew
  - Automation
- *Backs up* Automation
- *Troubleshoots*
- *Responds* to anomalies

AUTOMATION

- *Remembers*
  - Flight Plan
  - Altitude Restrictions
  - Procedures
- And *Reminds* the Pilot
- *Maintains* precise and efficient
  - Climbs and Descents
  - Flight Paths
  - Patterns
- And *Offers* them to the Pilot
- *Monitors* Aircraft and Pilot
- *Backs up* the Pilot
  - Tight and loose envelope warnings and protection
  - Danger avoidance with override

In summary

- Behind every successful AI, stands a strong human
  - *We are the enablers, the shepherds, and the turner-offers-and-turner-oners*

- A design that counts on humans to have machine strengths or machines to have human strengths is poor design
  - *I want to thank you for letting me be myself again*

- Human Machine Teaming is about the right allocation of functions and tasks between humans and machines – not just an easy interface and usability
  - *It is better to do good than to look good*

- Automation should comply with Crew Resource Management Principles
  - *Time to hold it to a higher standard*

- Metaphors can be useful in design, training, and usability
  - *Gonna fly like an Eagle*

- Sometimes the best design will give tasks to the human that the machine might do better, in order to improve the human's situation awareness
  - *Anything you can do, I can do better … but maybe you better handle this.*

A final question for you

- Where are the fully automated freight trains?
  - No passengers, Only 1.5D, Long, boring, dull trips. Perfect for full automation. And yet…
- Answers I don't buy
  - Unions – lots of countries don't have unions. Didn't save the conductors on the caboose!
  - Need to be able to stop to keep from hitting someone – they can't stop

Autonomous cars and aircraft should be no problem

Questions?

Adobe Firefly (Beta)
Image Not for Commercial Use

Backup Slides

Behind every successful AI, stands a strong human

- If you use computers in your job, chances are you will come across some behavior that might be described as **wonky,** at least once a month. Here are a few of my frequent visitors.
  - A spreadsheet cell won't highlight
  - You have to scroll up a page before it will let you scroll down
  - A file refuses to save
  - A forever spinning beach ball
  - A program just quits

- And yet, we manage to successfully complete the task
  - I assume the cells are selected and perform my duties without the reassurance
  - I scroll up before I scroll down so it will work – it's just this one document
  - I copy the contents of the file, paste it into a new file, and save that file
  - I turn it off and turn it back on again

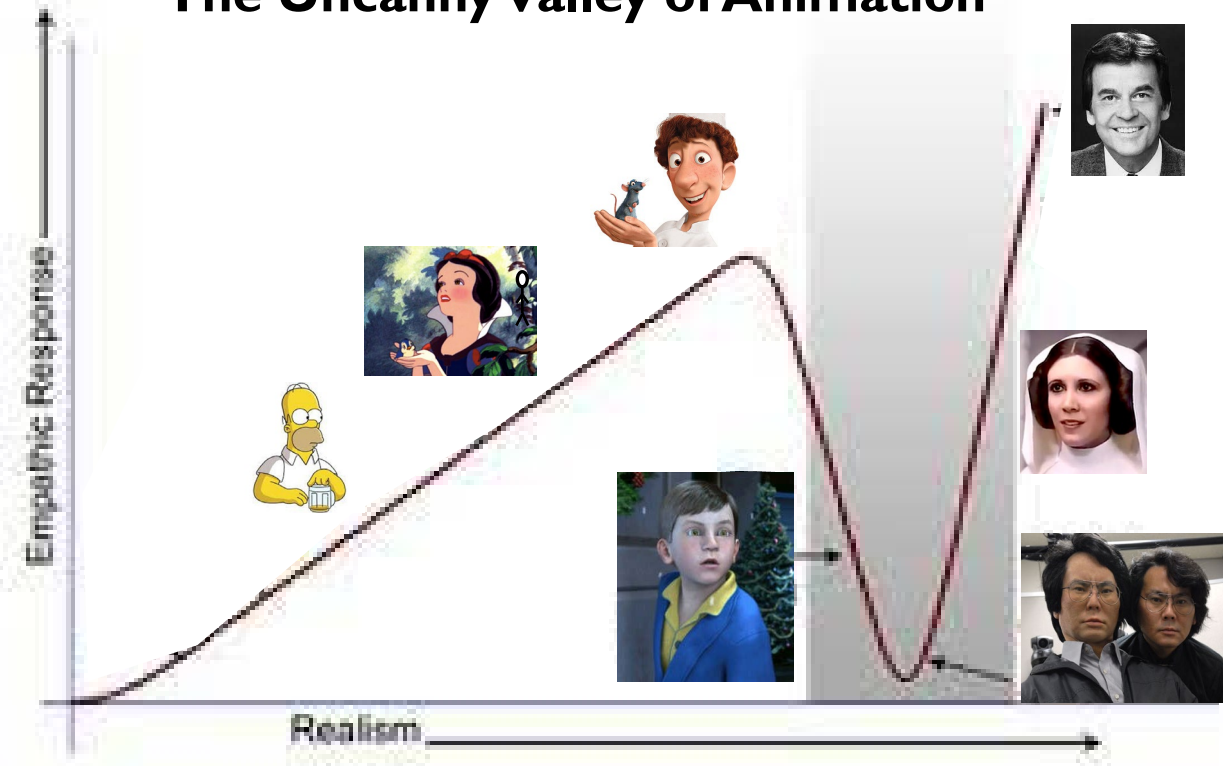- But a computer would stop under these conditions

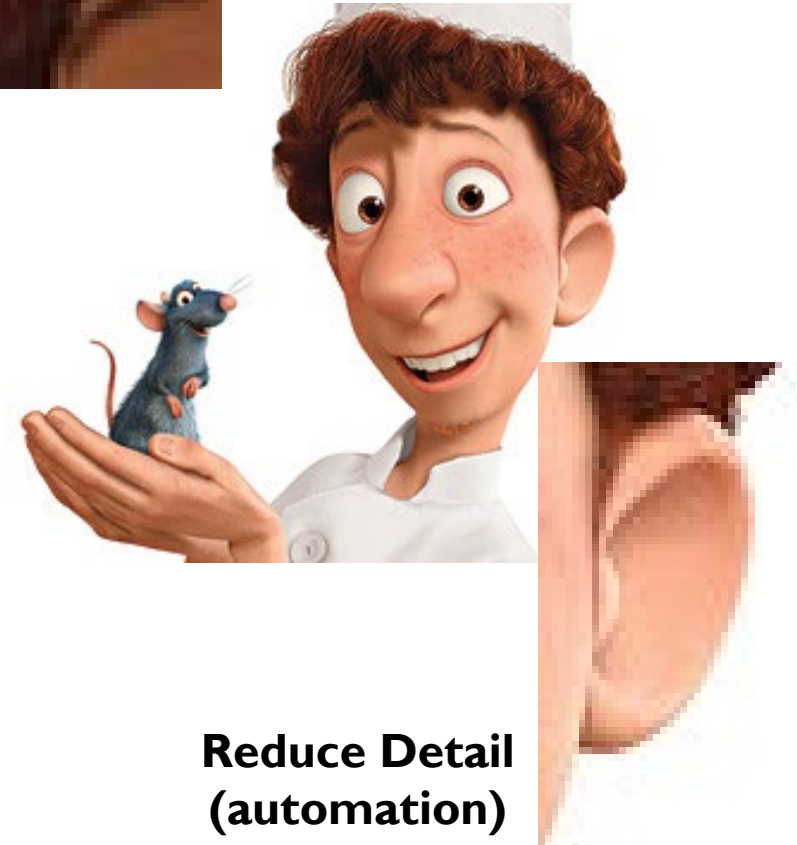# Humans won't be obsolete for a while
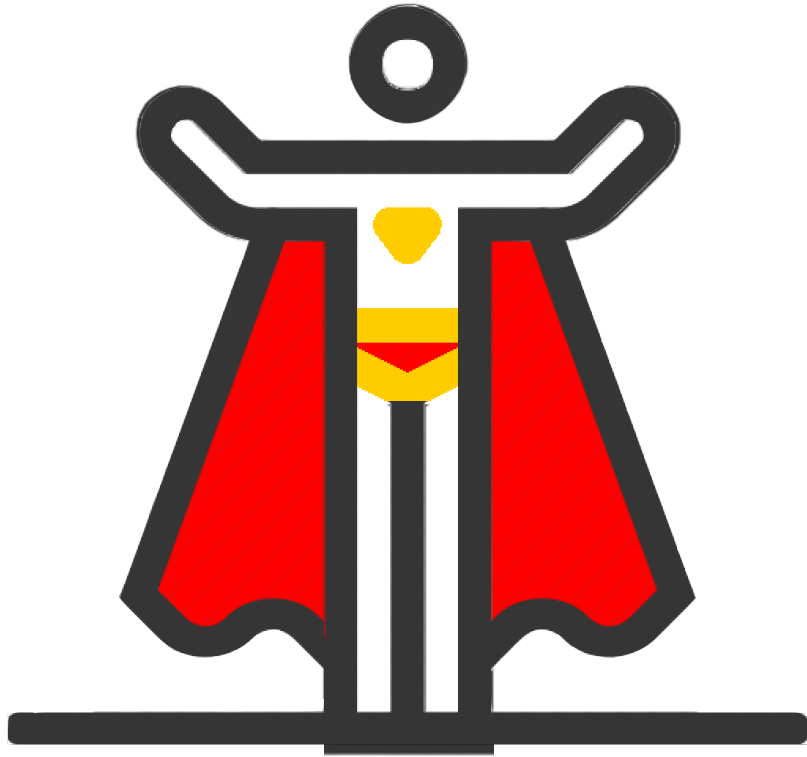
- Humans adapt
- Humans are goal driven
- Humans are multi-purpose
- Humans can be creative in unusual situations
- *Many systems would not work without humans to shepherd them across finish line*
- *We tame the bear.*

Take a lesson from PIXAR

**The Uncanny Valley of Animation**

**Lots of detail
(automation)**

**Reduce Detail
(automation)**

Replacing vs. Complementing

- The traditional view of Human-AI interaction is serial (e.g., levels of automation)
  - Human inputs/commands
  - Machine processes/performs
  - Human assess/completes

- The problem is that the world is dynamic and a particular 'level' of automation might be useful in one situation and disruptive in another
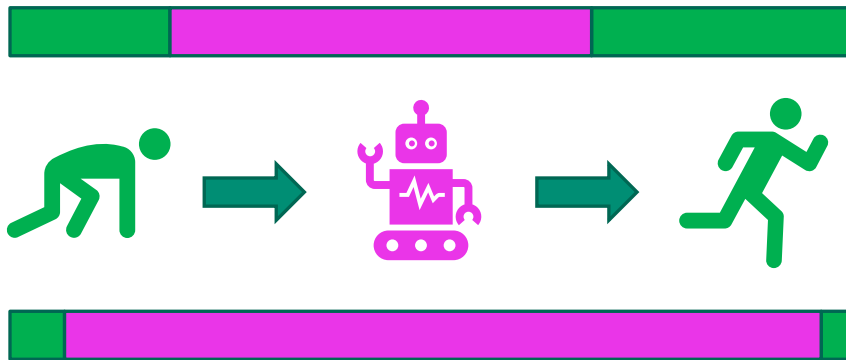

- We are researching a more collaborative interaction between humans and AI
  - Task switching / adaptable AI
  - Redundant tasking / processing
  - Mutual backup
  - Iterative
  - Automation as a better teammate
    - *Not necessarily human-human interaction*
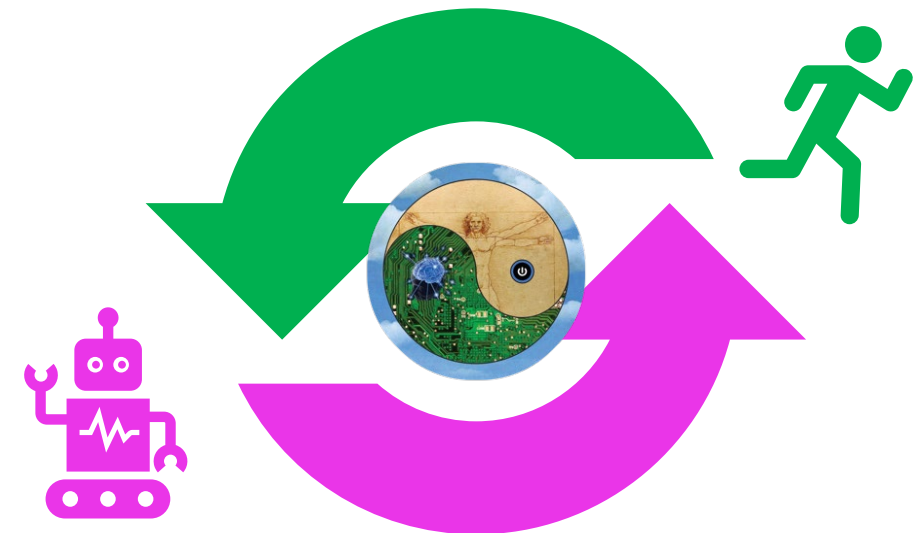
Foundation models are different animals

- Not necessarily rational
  - Not your grandfather's AI (e.g. mine)
    - Rule-based
    - Causal models
    - Logical deductions
    - Near-exhaustive searching
    - Predictable
  - Can be biased
  - Some are based on a subset of human behavior
  - Issues of data provenance, truth, predictability, and replicability

- Serious Trust Issues
  - Human-like behavior can produce undeserved over-trust
  - Explanations are extremely difficult to impossible
  - Can 'Hallucinate'
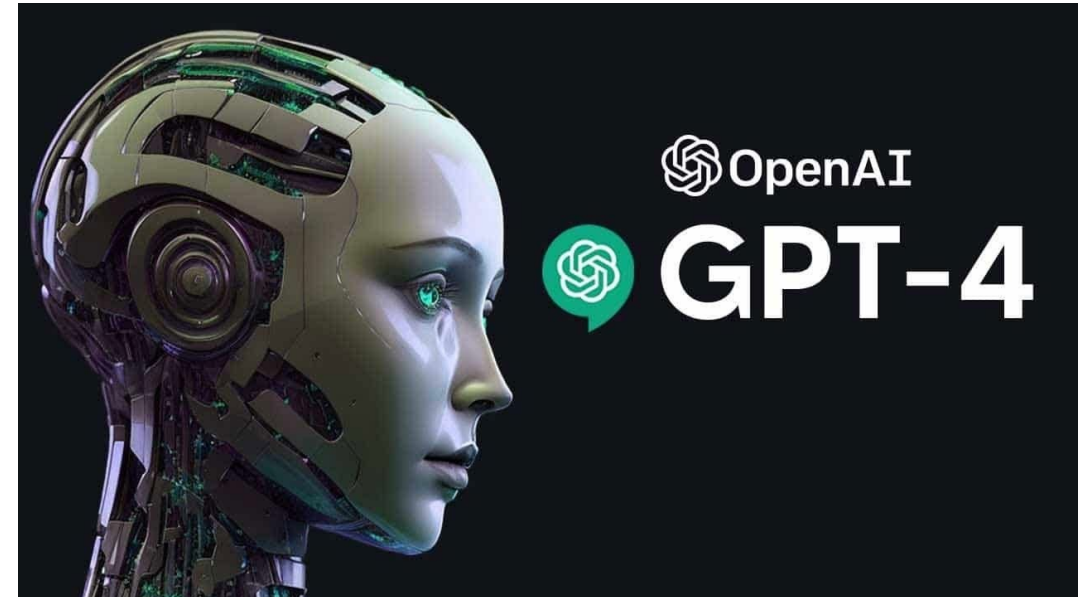
Massive ML models are not rational, they may be **much more robust**, but they may have a **greater risk for errors in judgement**.

Automation myths and misconceptions

Myth: More automation is always better, safer, cheaper.

- More automation is always better, safer, faster, cheaper.

Reality: There is no free lunch

- The more robust the automation, the more costly it can be to develop, manufacture, train and maintain.
  - Full automation nearly always comes with a cost or inconvenience
- Automation does fail – we turn it off and turn it back on again and it's fine.
- In most accidents, the automation has surpassed its operating thresholds and disengaged, leaving the human to handle the problem
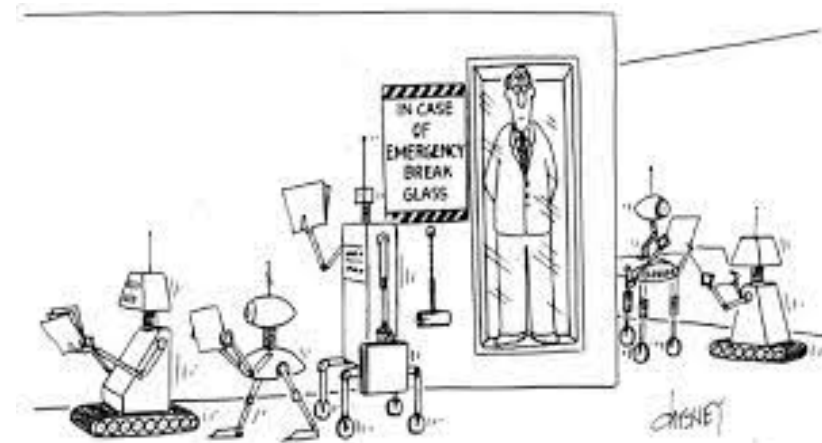
Automation myths and misconceptions

Myth: Full automation is the future, humans are slower, more expensive, and error-prone

- Humans are irrational and introduce errors.
  - 80% of all aviation accidents are caused by human error.

Reality: Automation is almost always limited/constrained with humans handling 'the rest.'

- Humans often save the day (so does automation) but there is little to no data collected on non-events.

- Situations often do not have time, data, or knowledge for fully rational decisions. Humans effectively use Naturalistic Decision Making to make effective decisions in these situations.

- Humans make bad software work.

KEEP CALM AND TURN IT OFF AND ON AGAIN

IN CASE OF EMERGENCY BREAK GLASS

Automation myths and misconceptions

Myth: I am a human, therefore I know human factors

Reality: We humans rarely know ourselves, let alone the entire population

- We've all used horrible interfaces and thought what idiot designed this? Well, there's a good chance that idiot had a Ph.D. and they definitely were human.

Launch Missile   **Coffee**