# Autonomous Cyber


Dr Dale A. Lambert PSM
Chief Cyber and Electronic Warfare Division
Defence Science and Technology

# Agenda

1. Cyber Reliance
2. Cyber Contest
3. Autonomous Cyber

# 1. Cyber Reliance

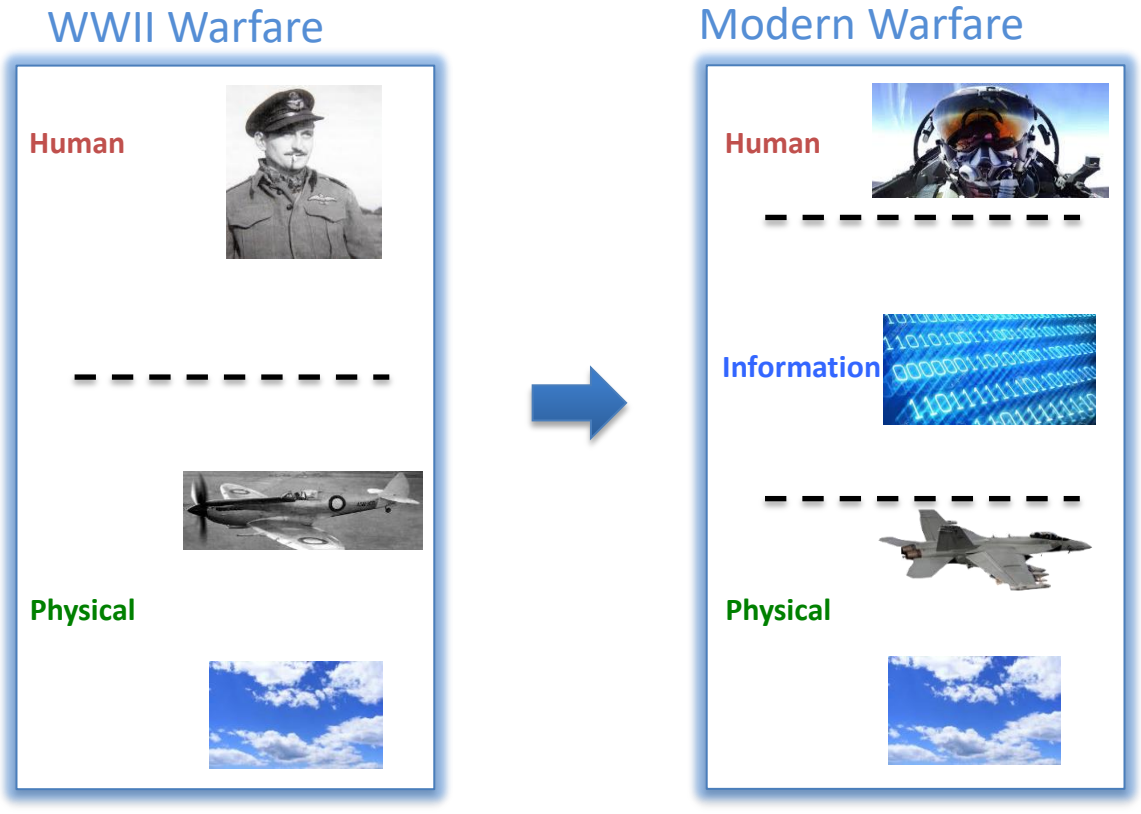
Society is now totally reliant on the information environment.

# From the Industrial Age to the Information Age

## Narrative

- The Information Revolution has transformed warfare.
- WWII represents Industrial Age Warfare – people interact directly with their industrial machines to fight the war.
- Modern Warfare is characterised by Information Age Warfare – people now interface to an information environment that in turn interfaces to their industrial machines.

## WWII to Modern Warfare



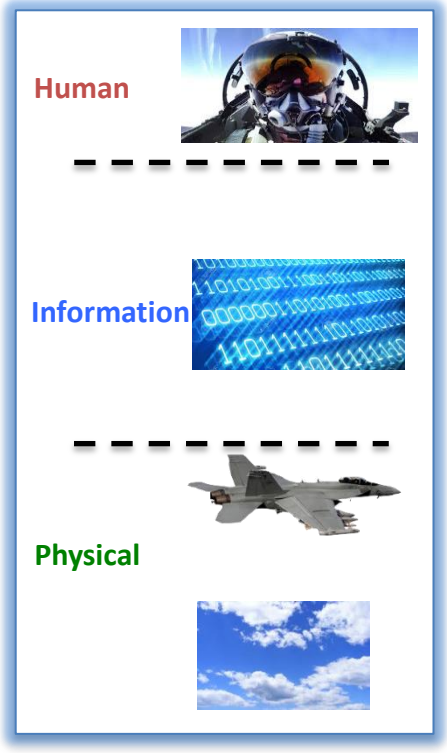
# Cyber Reliance

## Narrative

- The Information Revolution has had a ubiquitous effect on society, not just warfare.
- The physical and human environments in society now have a *reliance* on the information environment that separates them.
- *We are constructing a digital representation of the world and making it accessible to anyone, anywhere.*

## Cyber Reliance

- **Human**: we are describing almost everything about humanity digitally – of the 7.6 billion people in the world, there are over 2 billion Facebook users and over 1 billion WeChat users.
- **Information**: we are distributing almost everything digitally, connecting all our physical and virtual devices into the internet of things.
- **Physical**: we are sensing almost everything in the physical world digitally - there will be 1 trillion individual sensors in the world by the early 2020s.



# Physical Reliance on the Information Environment

## Narrative

- Our physical industrial world is now reliant on the information environment.
- This includes society's critical infrastructure.

## Cyber Reliance

**Power and Water**

distribution via Supervisory Control and Data Acquisition (SCADA) systems

**Telecommunications**

SCADA, software defined networking dominating telecommunications

**Agriculture**

tractors with computers that are automatically adjusting fertilizers

**Hospitals**

computer controlled patient medication and records

**Transportation**

automated trains and semi-autonomous cars

**Security**

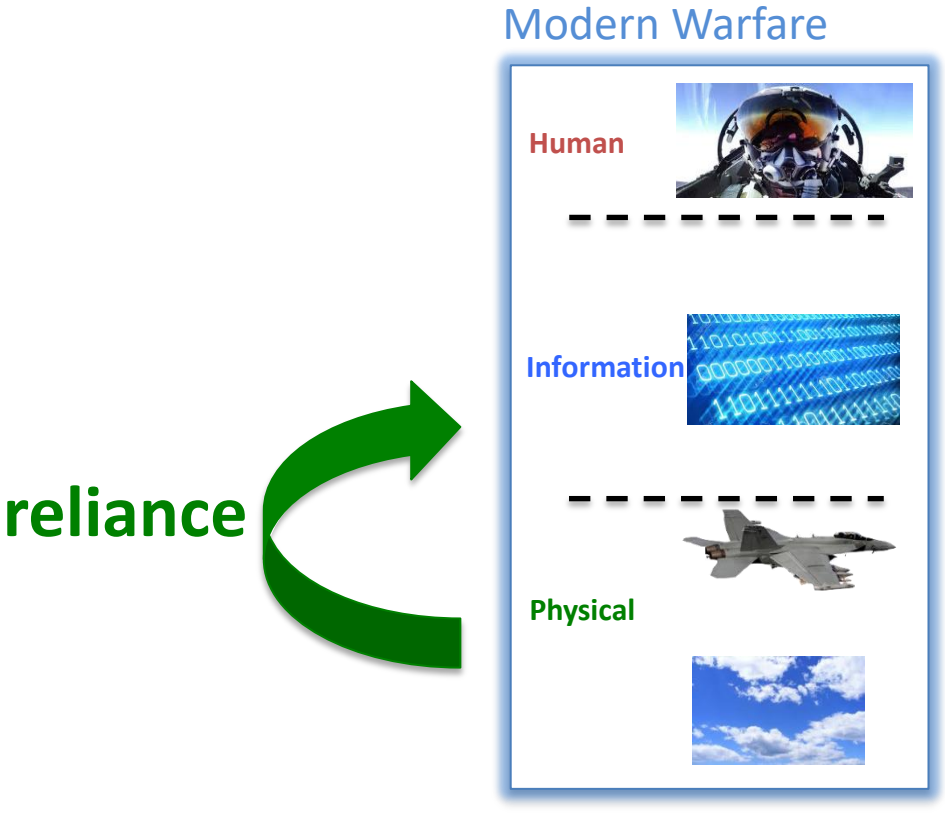
information controlled Defence systems

# Physical Reliance on the Information Environment

## Conclusion

- People no longer directly control the physical industrial world.
- People instead issue commands to an information environment and that information environment directly controls the physical industrial world.
- *If you control the information environment, you control the physical industrial world!*

## Cyber Reliance



# Human Reliance on the Information Environment

*What is truth?*

*There are two dominant theories of truth*

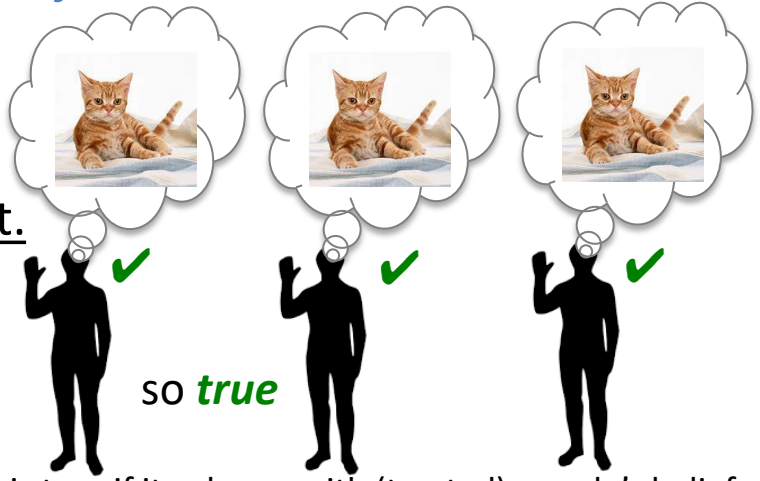
a cat ✔

a mat ✔

cat on mat ✔

so ***true***

e.g. The cat is on the mat.



so ***true***

The statement is true if it corresponds with the world.

The statement is true if it coheres with (trusted) people's beliefs.

## The Correspondence Theory of Truth

## The Coherence Theory of Truth

*live streaming of the Moscow weather through Earthcam.*

*But both approaches to truth are now decided through screens!*

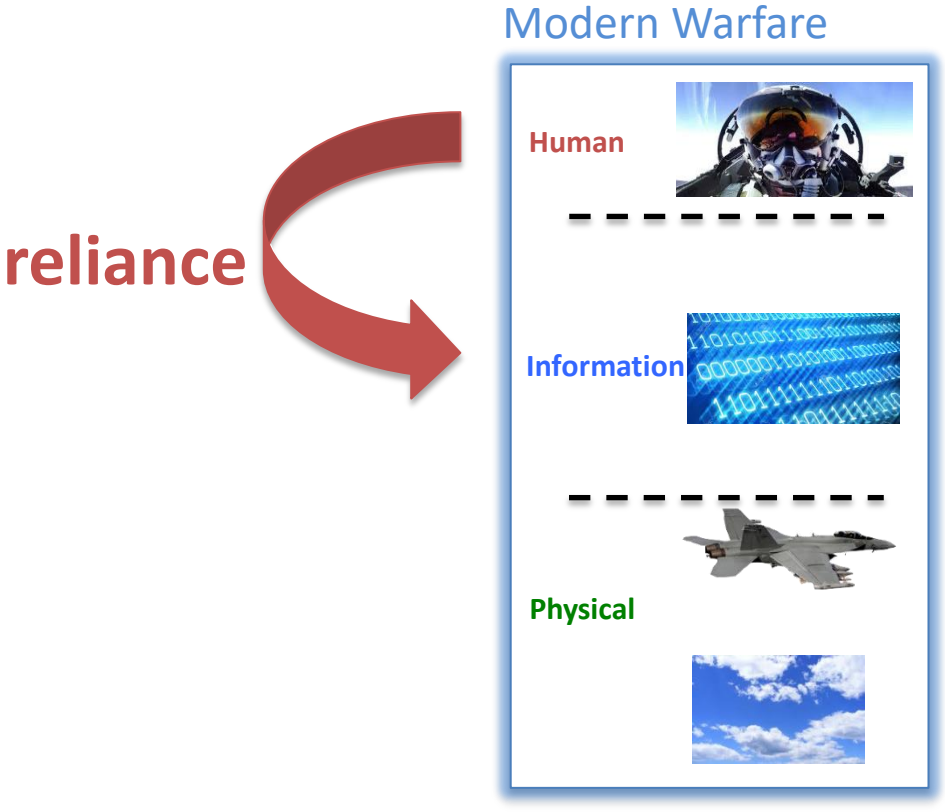

*social media, e.g. Facebook and Twitter.*

# Human Reliance on the Information Environment

## Conclusion

- The Correspondence Theory of Truth assigns truth based on correspondence with the world, but this now comes through digital images of the world *that can be manipulated by image and video editors*.
- The Coherence Theory of Truth assigns truth based on coherence of opinion, but this now comes through social media *that can be manipulated by fake news*.
- *If you control the information environment, you control peoples' truth!*

## Cyber Reliance



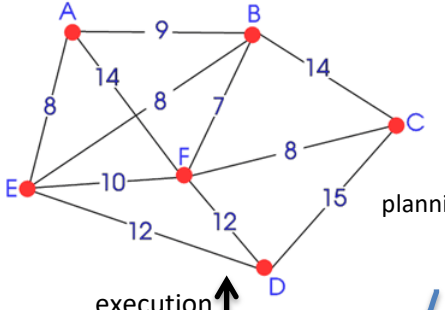
# Cyber Reliance on the Information Environment

## e.g. algorithmic warfare: attacking and defending algorithms in the information environment.

### Features not just flaws

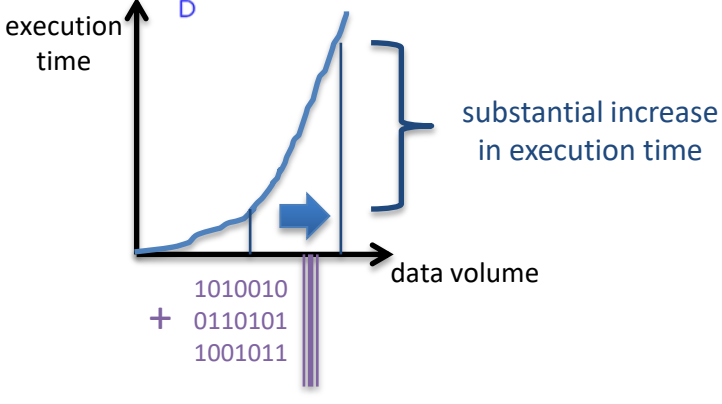
- There is a tendency to *only* focus on flaws in the coding of algorithms and believe that algorithmic warfare threats are eliminated once the code is appropriately patched.
- This is <u>not</u> the case. Algorithmic warfare threats can come from the *features* of *perfectly* coded algorithms, not just *flaws* in coded algorithms.

1. A *perfect* travelling salesman algorithm can be defeated by simple data overload because it provably requires exponential time to compute.




**<u>Travelling Salesman Problem</u>:**
**Visit every object while minimising total traversal cost.**

planning, logistics, microchip manufacture, DNA sequencing

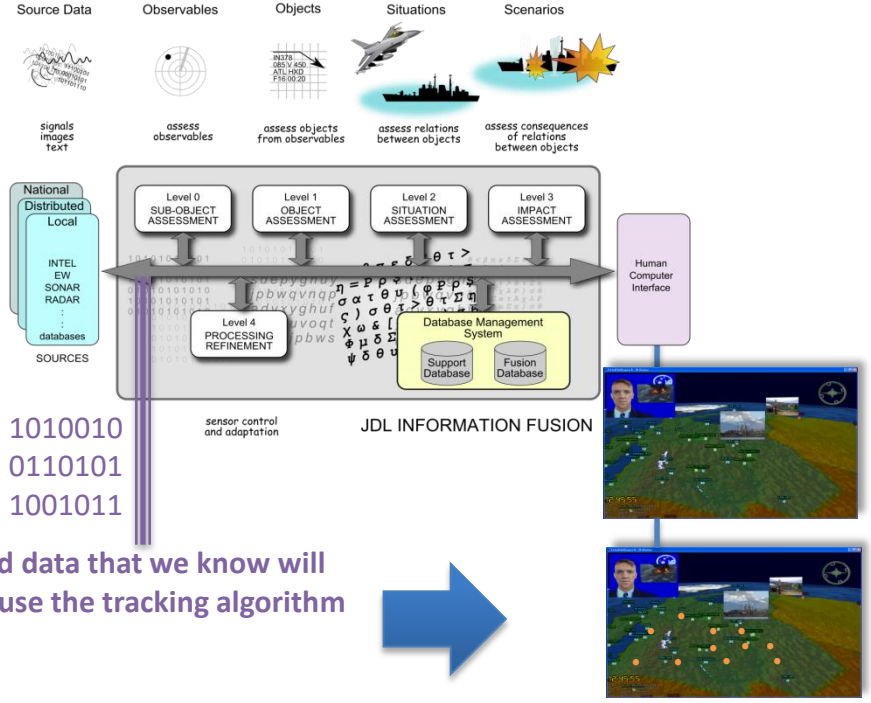



**Simple Data Overload**

# Cyber Reliance on the Information Environment

## e.g. algorithmic warfare: attacking and defending algorithms in the information environment.

2. A *perfect* tracking algorithm can be defeated by inserting *specific* data that causes the algorithm to create false tracks and lose true tracks.




3. Since 2011, machine learning image recognition algorithms have significantly outperformed people, but *perfect* machine learning algorithms are also subject to algorithmic warfare.

**Simple Image Perturbation**

1010010
0110101
1001011

machine classified image as a school bus → 10x stretch small perturbation → machine misclassified image as an ostrich

# Cyber Reliance on the Information Environment

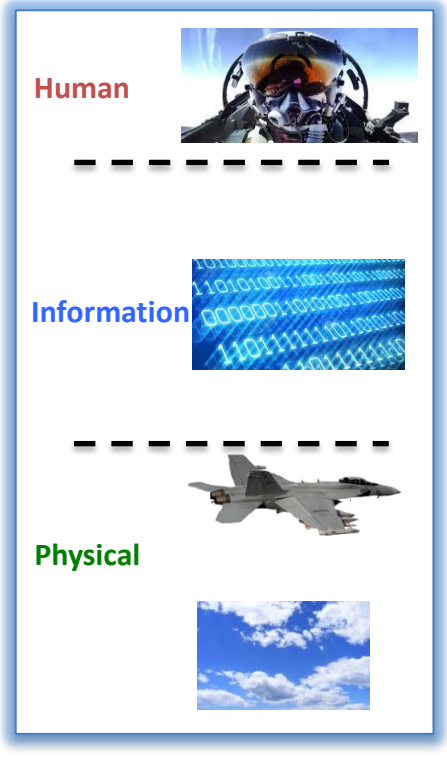
## Conclusion

- The power of the information environment lies in its algorithms.
- Flawed algorithms provide vulnerabilities that can be exploited until patched.
- But perfect algorithms can also provide vulnerabilities because they have features that can be exploited.
- *If you control the information environment (algorithms), then you control the information!*

## Cyber Reliance

**reliance**

Modern Warfare

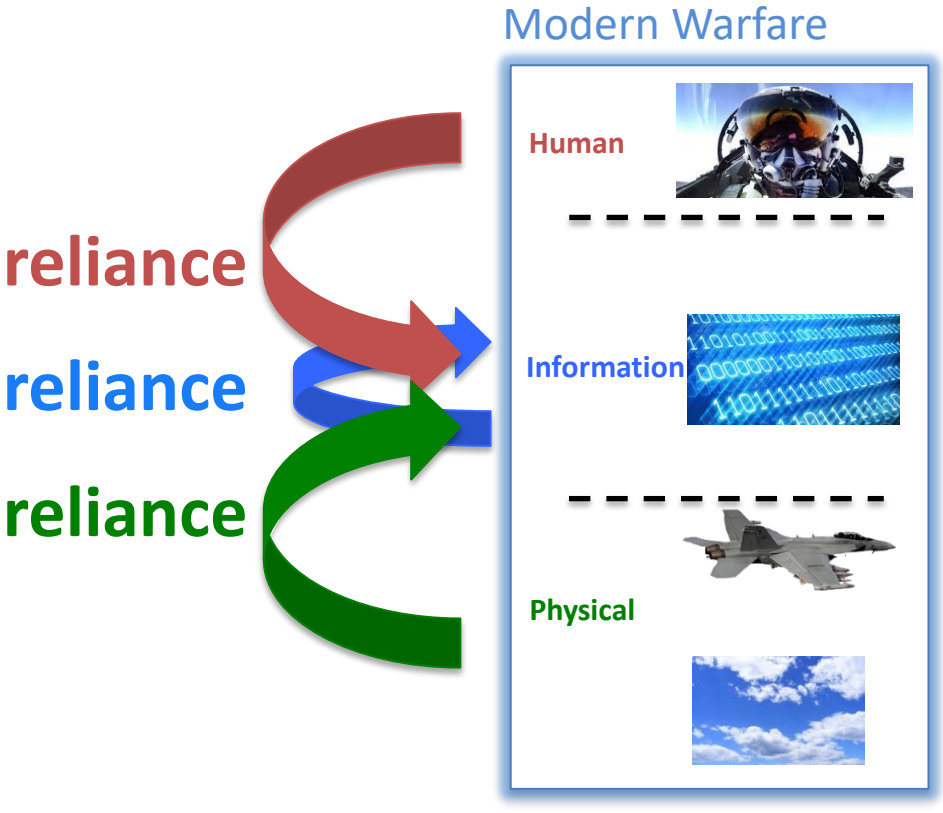


# Cyber Reliance

## Conclusion

- *If you control the information environment, then you control the physical industrial world!*
- *If you control the information environment, then you control peoples' truth!*
- *If you control the information environment (algorithms), then you control the information!*
- ***If you control of the information environment, you have control!***

## Cyber Reliance



# 2. Cyber Contest

Fighting the war through and inside of the information environment.

# Cyber Contest

## Information Environment Battlefield

The “weaponisation” of data, ideas and goods by adversaries such as Russia poses a bigger risk than “missiles and tanks”.

“The emergence of these disruptive technologies is so prevalent and rapid that what we need now is an urgent reappraisal of how, with what and by whom war is waged in the future.”

– General Sir Mark Carleton-Smith,
UK Chief of Army,
Jun-19.

- ***If you control of the information environment, you have control!***

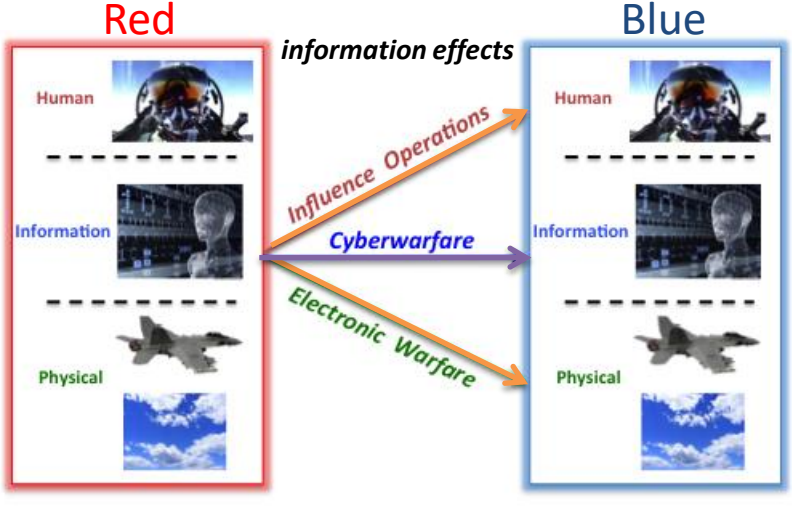
## Cyber Contest

1. ***fight the war through the information environment***

information environment: primary conduit for effects outside the information environment because it offers access and scale.

2. ***fight the war inside of the information environment***

information environment: fight the war inside the contested information environment because it is the new battlefield.

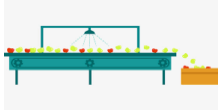


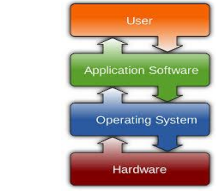
# Lambert's IW Maturity Model

*australian industry* *cyber industry*

**Design**


**Implementation**






**Proof**

**Testing**



**Secure**

protect

*against* **Penetration**

attack

***development***

***- develop capability to provide function***

***✗ function errors***

***verification***

***- develop capability to provide safe and reliable function.***

***✗ function attacks***

***security***

***- develop capability to provide safe secure and reliable function.***

***✗ wall breaches!!!***

- Cyber security needs automation to find malicious needles in normal haystacks.
- Cyber security needs automation to handle the speed of cyber operations.
- Cyber security needs automation to compensate for a large personnel deficit.

- There is significant commercial effort to apply AI to Cyber security.
- Current AI systems significantly outperform people at speech processing, image processing and at most games.
- Cyber Grand Challenge in 2016: machines that automatically patch vulnerabilities in own operating systems, craft exploits for other systems.

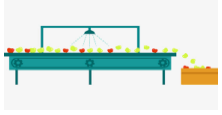
***information environment***

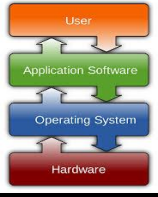
- In 1949 Oxford philosopher Gilbert Ryle coined the phrase "ghost in the machine" for the mind.
- AI systems are the actual ghosts in the machine.

# Lambert's IW Maturity Model

**Design**


**Implementation**






**Proof**

**Testing**



**Secure**

protect

*against* **Penetration**

attack

**Contest**

attack

defend

protect

*requires* **Aware**

shape

analyse

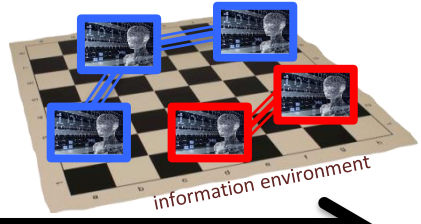
sense

**Coordinate**

command & control

*requires* **Communication**

communicate




***development***

***- develop capability to provide function***

***✗ function errors***

***verification***

***- develop capability to provide safe and reliable function.***

***✗ function attacks***

***security***

***- develop capability to provide safe secure and reliable function.***

***✗ wall breaches***

***AI digital ghosts***

***- develop capability that can contest to provide safe, secure and reliable function.***

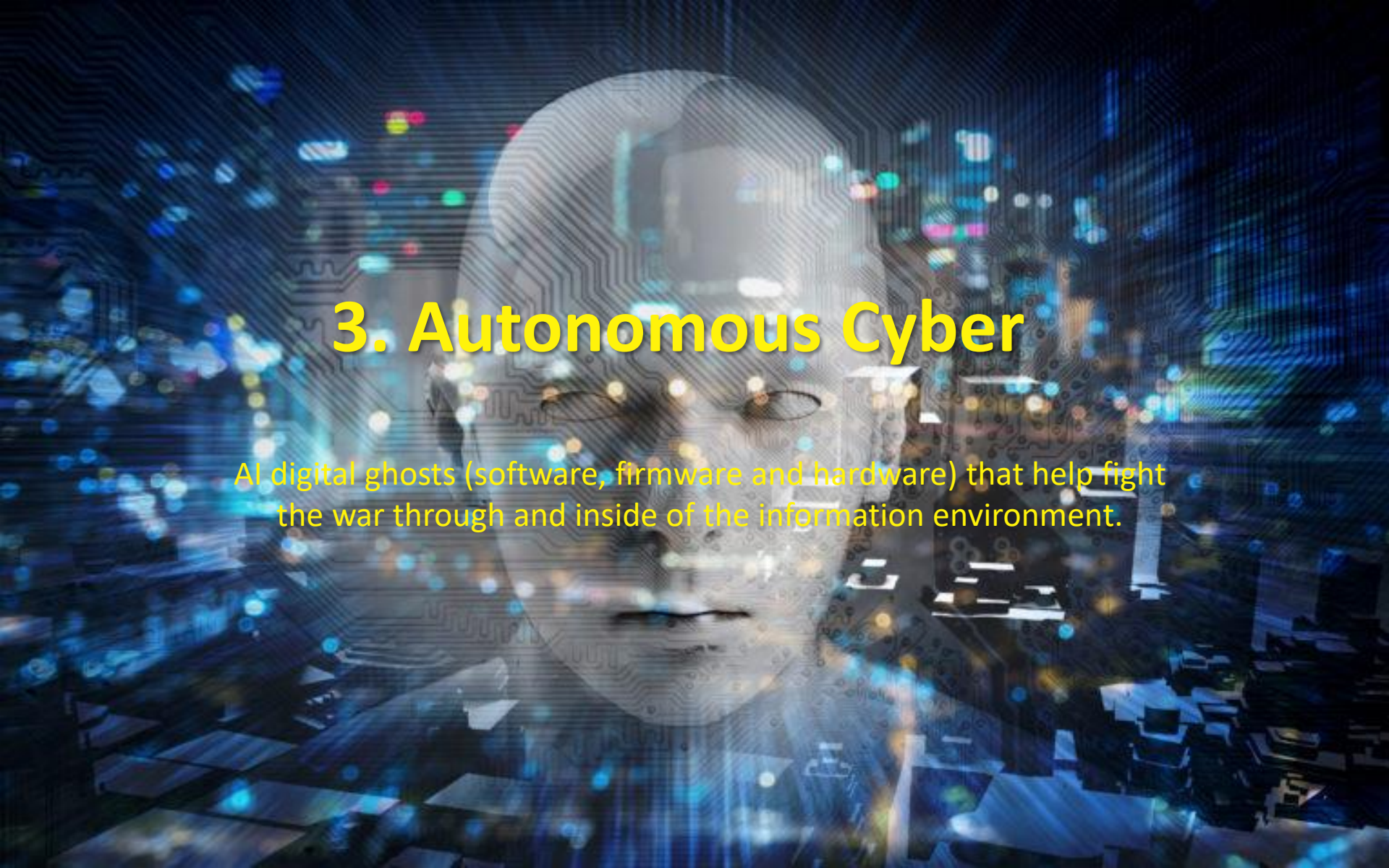
***✗ isolated intents***

***AI digital force***

***- develop coordinated capability that can contest to provide safe, secure and reliable function.***

***What is AI's capacity to deliver?***

***maturity***


3. Autonomous Cyber
AI digital ghosts (software, firmware and hardware) that help fight the war through and inside of the information environment.

# Artificial Intelligence

**Search Phase**

## Problem Solving

Intelligence was conceived narrowly as the ability to solve problems.

**Winning at games was a typical emphasis.**

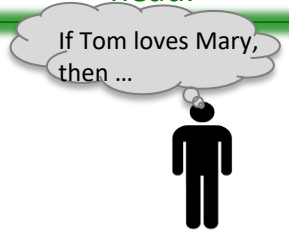
## Solipsism

Intelligence was about thinking inside the machine.

**AI machines operated in their own mental bubble, disconnected from the physical world.**

## Manipulate Symbols

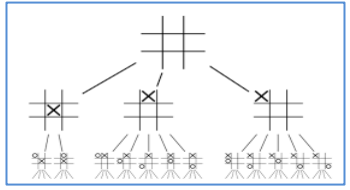
Thinking was about manipulating symbols in the head.




**Programs manipulate symbols inside the machine with users connecting those symbols to the outside world.**

## Trial-Error Search

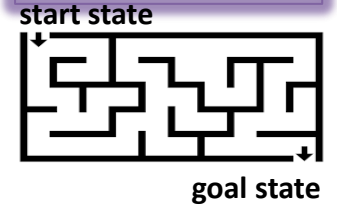
Thinking involved trial and error search.



**Programs executed trial and error search by manipulating symbols inside the machine.**

## General Purpose Search

A general search method was the basis of all problem solving.




**A general purpose search program could solve any problem when given the start state, goal state and data for that problem.**

# Artificial Intelligence

**Search Phase**

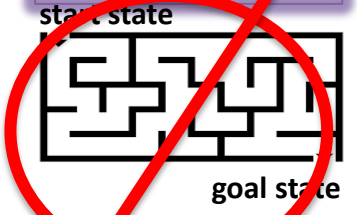

"We must hypothesize from our experience to date that the problem solving power exhibited in an intelligent agent's performance is primarily a consequence of the specialist's knowledge employed by the agent, and only very secondarily related to the generality and power of the inference method employed. Our agents must be knowledge-rich, even if they are methods-poor."

- Feigenbaum (1977)

"But we have agreed to set aside the problem of acquiring knowledge till we better understand how to represent and use it."

- Minsky (1968)

A general search process was the basis of all problem solving.

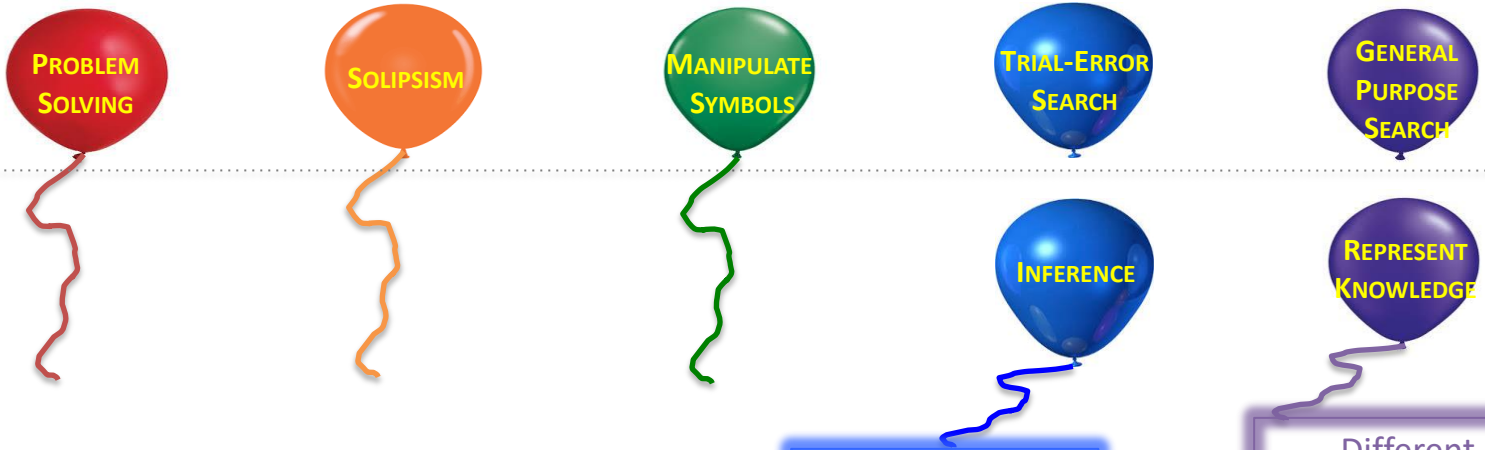



A general purpose search program could solve any problem when given the start state, goal state and data for that problem.

# Artificial Intelligence

**Search Phase**

**Representation Phase**




Search became inference.

Different knowledge representation schemes arose.

✓ *Knowledge representation became good enough to support a shift in emphasis to knowledge content.*

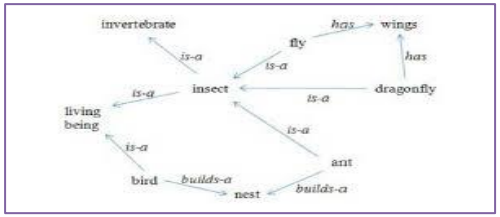
```
male(dale). parent(dale, adrian).
father(X, Y) :- parent(X, Y), male(X).
child(X, Y) :- parent(Y, X).
```

**Logic Programming combined declarative logic and procedural search**

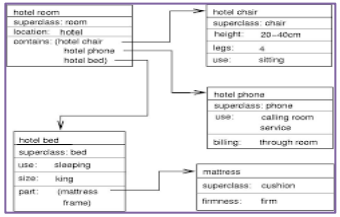
**Debate between declarative knowledge (facts) and procedural knowledge (recipes)**

**Formal logic representations held sway over time.**

**Some examples:**



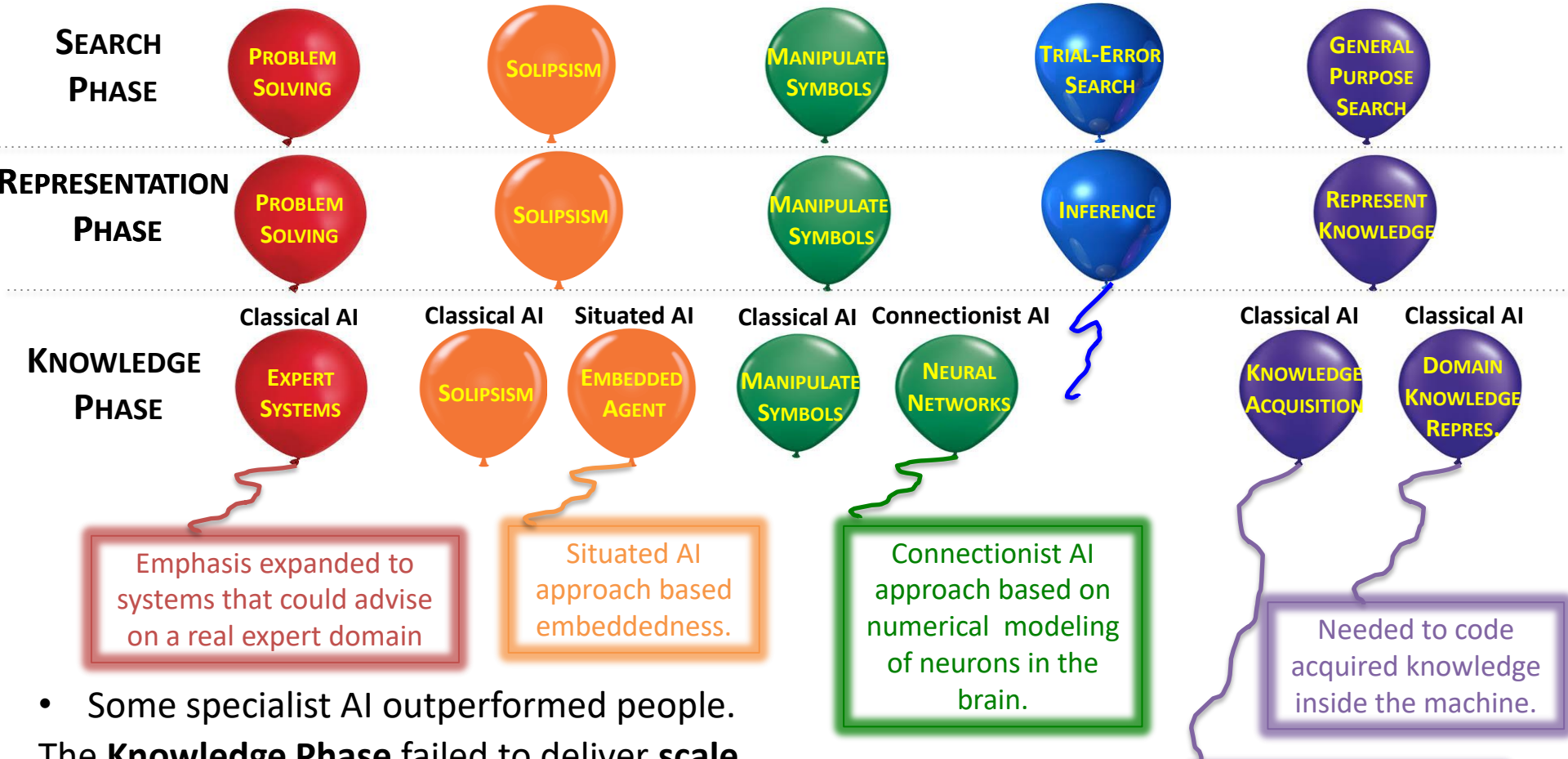

semantic networks




frames


formal logics

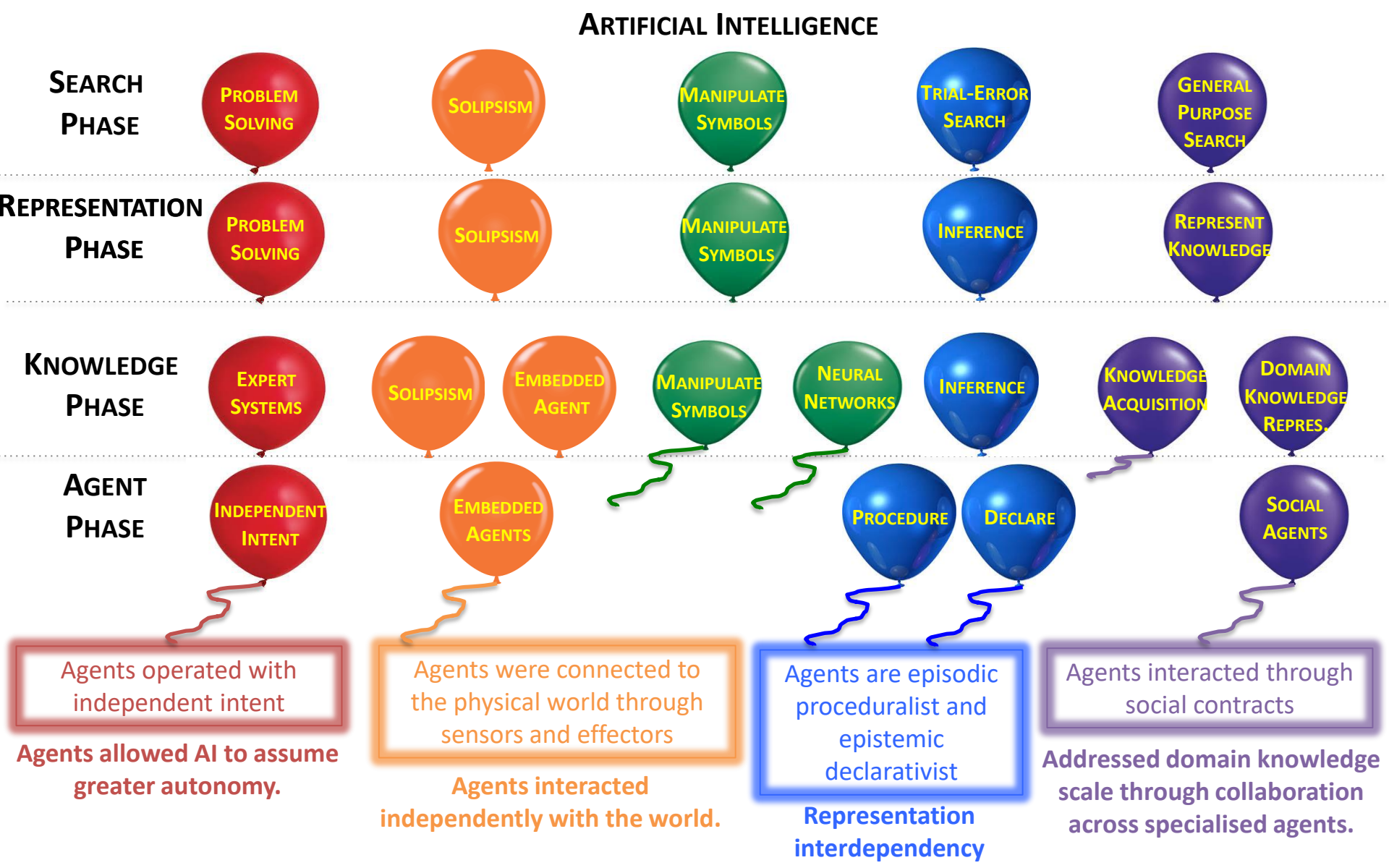
# Artificial Intelligence



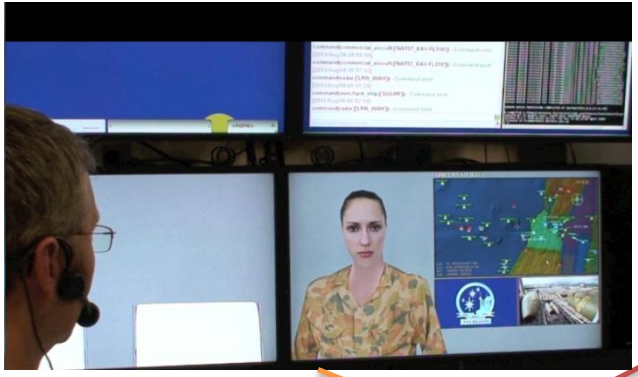

- Some specialist AI outperformed people.

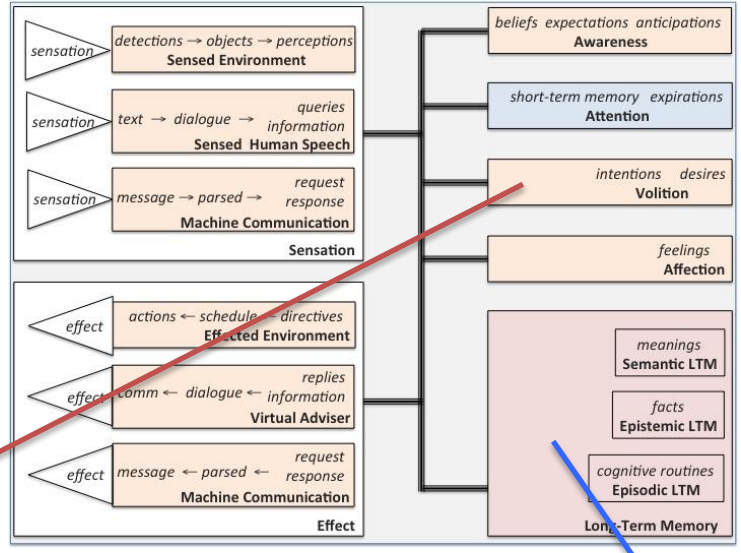
The **Knowledge Phase** failed to deliver **<u>scale</u>**.

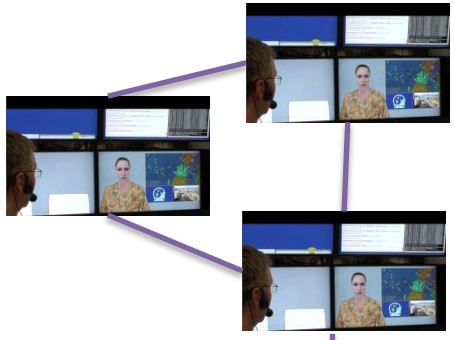
- Classical AI was brittle without *scale*, and unwieldy with *scale*.
- Connectionist AI did not *scale* beyond simple classification problems.
- Situated AI did not *scale* to cognitive functions beyond the present.


ARTIFICIAL INTELLIGENCE
SEARCH PHASE
PROBLEM SOLVING
SOLIPSISM
MANIPULATE SYMBOLS
TRIAL-ERROR SEARCH
GENERAL PURPOSE SEARCH
REPRESENTATION PHASE
PROBLEM SOLVING
SOLIPSISM
MANIPULATE SYMBOLS
INFERENCE
REPRESENT KNOWLEDGE
KNOWLEDGE PHASE
EXPERT SYSTEMS
SOLIPSISM
EMBEDDED AGENT
MANIPULATE SYMBOLS
NEURAL NETWORKS
INFERENCE
KNOWLEDGE ACQUISITION
DOMAIN KNOWLEDGE REPRES.
AGENT PHASE
INDEPENDENT INTENT
EMBEDDED AGENTS
PROCEDURE
DECLARE
SOCIAL AGENTS
Agents operated with independent intent
Agents allowed AI to assume greater autonomy.
Agents were connected to the physical world through sensors and effectors
Agents interacted independently with the world.
Agents are episodic proceduralist and epistemic declarativist
Representation interdependency
Agents interacted through social contracts
Addressed domain knowledge scale through collaboration across specialised agents.

# Artificial Intelligence





**Agent Phase**

**Independent Intent**

Agents operated with independent intent

**Agents allowed AI to assume greater autonomy.**

**Embedded Agents**

Agents were connected to the physical world through sensors and effectors

**Agents interacted independently with the world.**

**Procedure** **Declare**

Agents are episodic proceduralist and epistemic declarativist

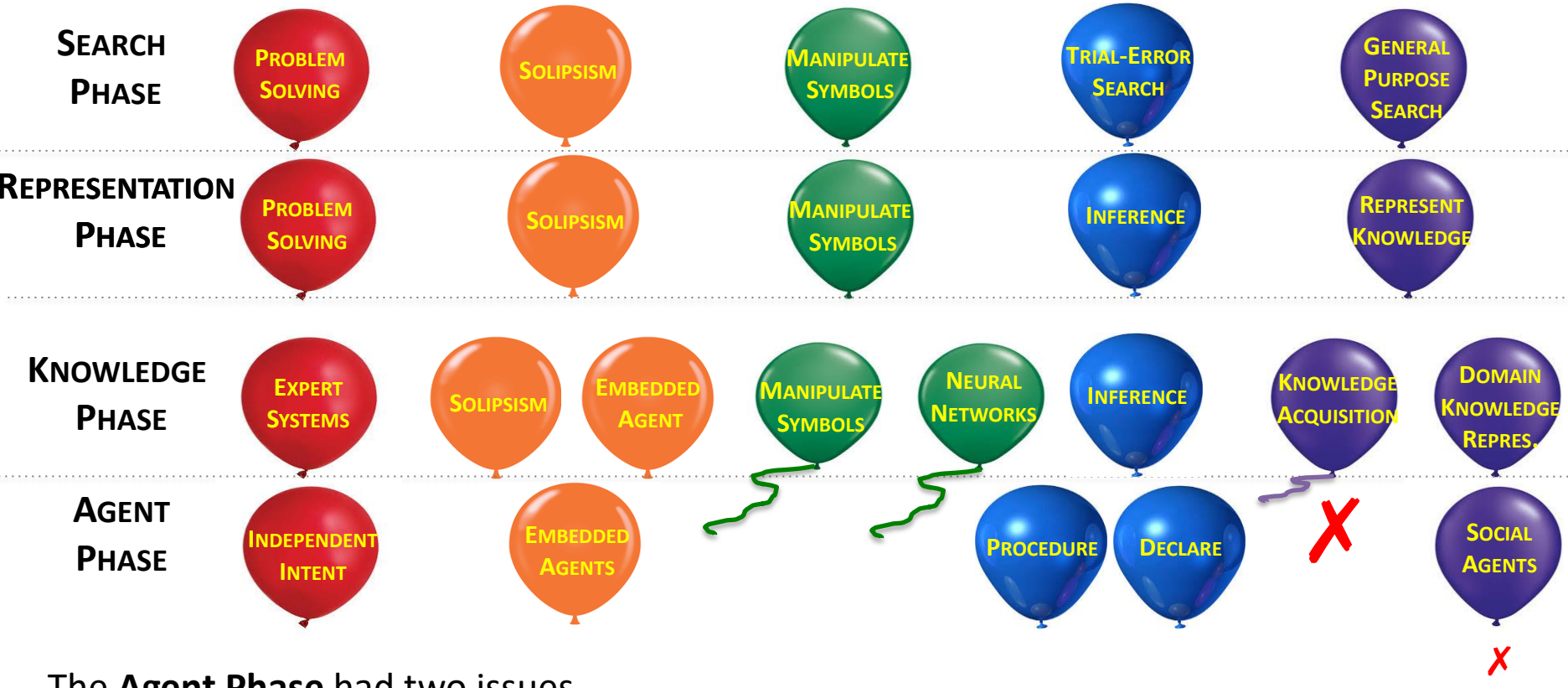
**Representation interdependency**

**Social Agents**

Agents interacted through social contracts

**Addressed domain knowledge scale through collaboration across specialised agents.**

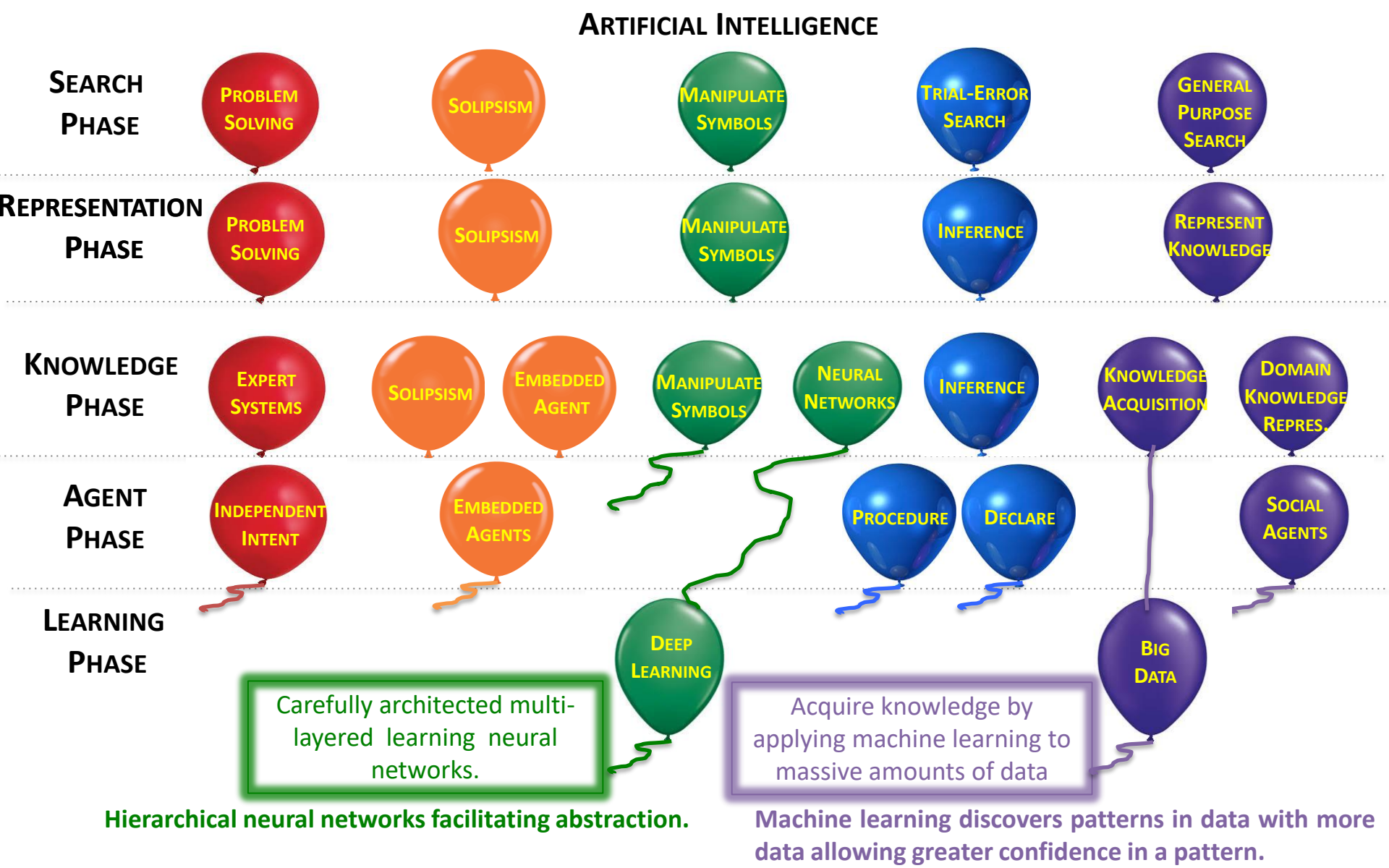
# Artificial Intelligence




The **Agent Phase** had two issues.

- Knowledge Acquisition remained a tedious hand coded exercise.
- The Agent Phase confronts a tradeoff between optimality and decentralised scale.

# Artificial Intelligence




**Hierarchical neural networks facilitating abstraction.**

**Machine learning discovers patterns in data with more data allowing greater confidence in a pattern.**

# Learning Phase

## Connectionist AI Success

- Deep Learning has produced outstanding success in:
  - image processing;
  - speech processing; and
  - language translation,

  significantly out performing people since around 2011.
- As a standout example, AlphaGo demonstrates the success of Deep Learning at the board game Go.
- AlphaZero (2017) extended to Chess and beat the world's best chess program.
- "It doesn't play like a human, and it doesn't play like a program … It plays in a third, almost alien, way."

## AlphaGo

- AlphaGo uses tree search advised by CNN deep learning heuristic predictions. [Search Phase + Learning Phase]
- It was trained using a database of 30 million examples.
- In 2015 AlphaGo was the first program to beat a professional Go player.
- In 2016 AlphaGo was the first program to beat a 9-dan Go professional, winning 4 out of 5 games.
- In 2017 AlphaGo beat the world number 1 ranked Go player, winning 3 out of 3 games.

# Learning Phase Issues

## Deep Learning Issues

**Abstract Concepts**: deep learning is not learning abstract semantic concepts.

"This sensitivity to adversarial examples suggests that these CNNs are not learning semantic concepts in the dataset. … To this end, we posed our main hypothesis: *the current incarnation of deep neural networks have a tendency to learn surface statistical regularities as opposed to high level abstractions*".

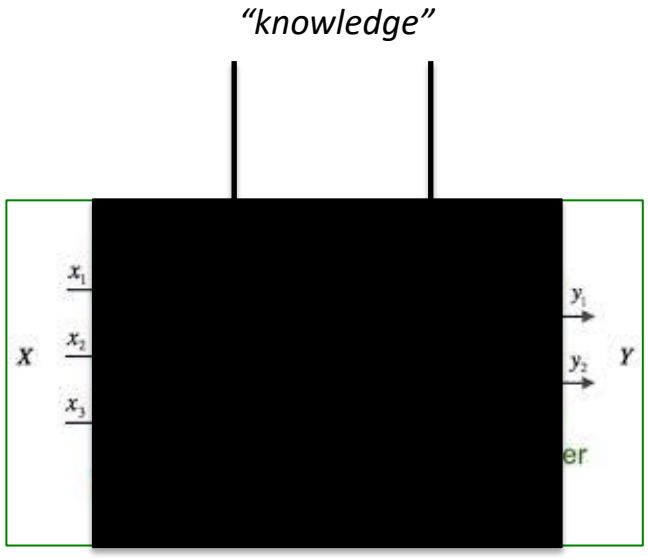
- Jo and Bengio (2017) Bengio, LeCun, Bottou and Haffner (1998) invented CNNs.

# Learning Phase Issues

## Deep Learning Issues

**Explanation and Trust**: deep learning "knowledge" is stored as numerical weights between nodes and not symbolically.

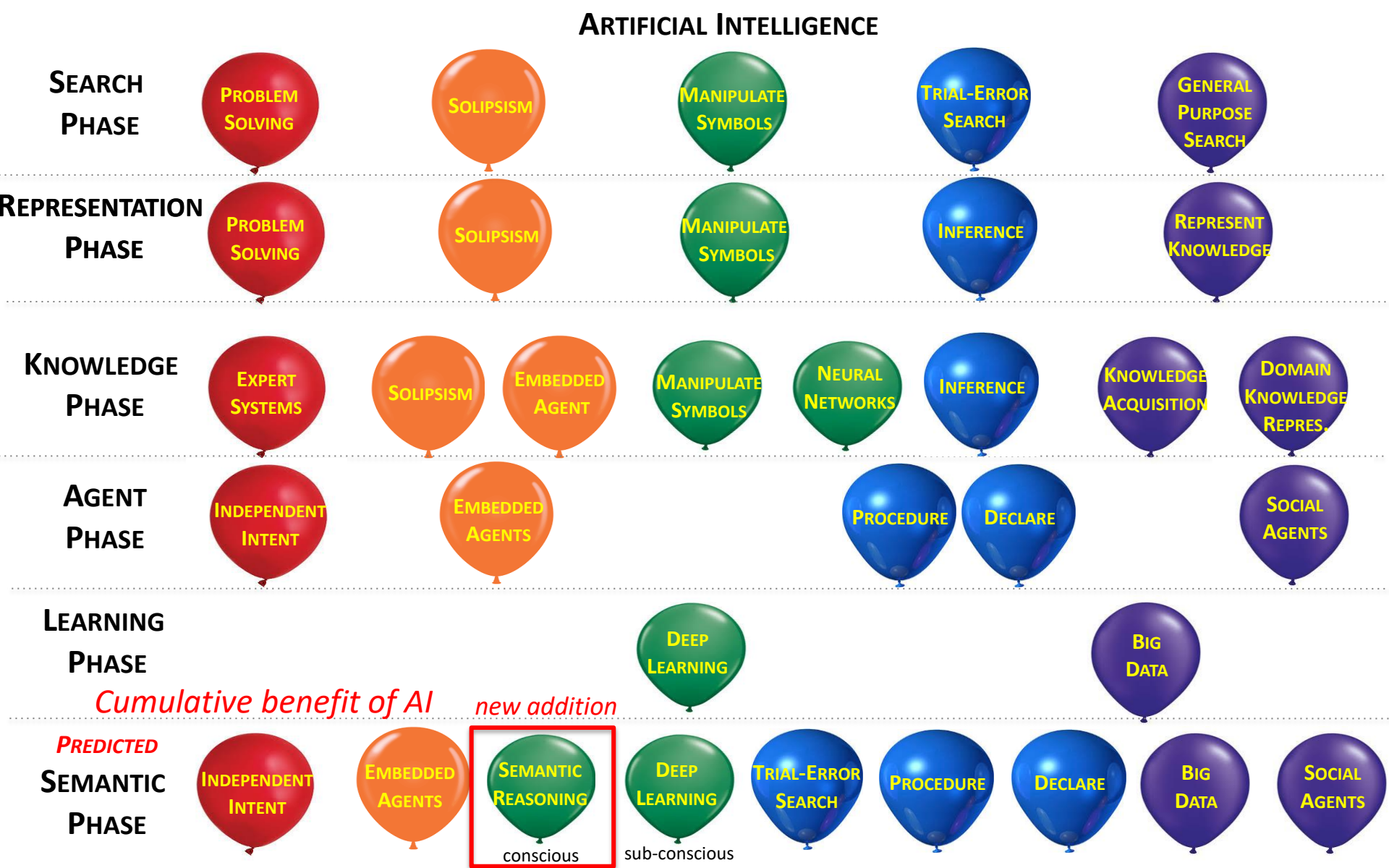
- *therefore*, the system cannot symbolically **explain** its choice of output for each given input.
- *therefore*, the system may not engender human **trust**.




**It is a black box and it plays like an alien!**

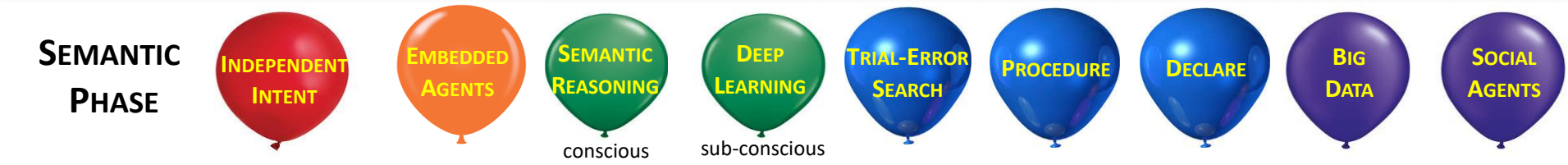
- *There is significant commercial effort to applying machine learning to cyber security.*
- *But if our AI digital ghosts are going to fight the war inside the machines then we want to understand what they are doing and why.*

# Artificial Intelligence



Conclusions:

1. Society is now totally reliant on the information environment.
2. We need to fight the war through and inside of the contested information environment.
3. We need trusted AI digital ghosts to successfully fight.
4. The Semantic Phase of AI is needed to deliver our digital ghosts.



# Questions?


Dr Dale A. Lambert PSM
Chief Cyber and Electronic Warfare Division
Defence Science and Technology
dale.lambert@dst.defence.gov.au
+61 8 7389 6612