Challenges in Engineering Responsible Technology – Towards Ethical AI

Dr Bertie Müller

University of South Wales & AISB

bertie.muller@southwales.ac.uk

University of
South Wales
Prifysgol
De Cymru

Overview

01// Engineering MAS

Before the current resurgence of AI …

03// The AI Hype

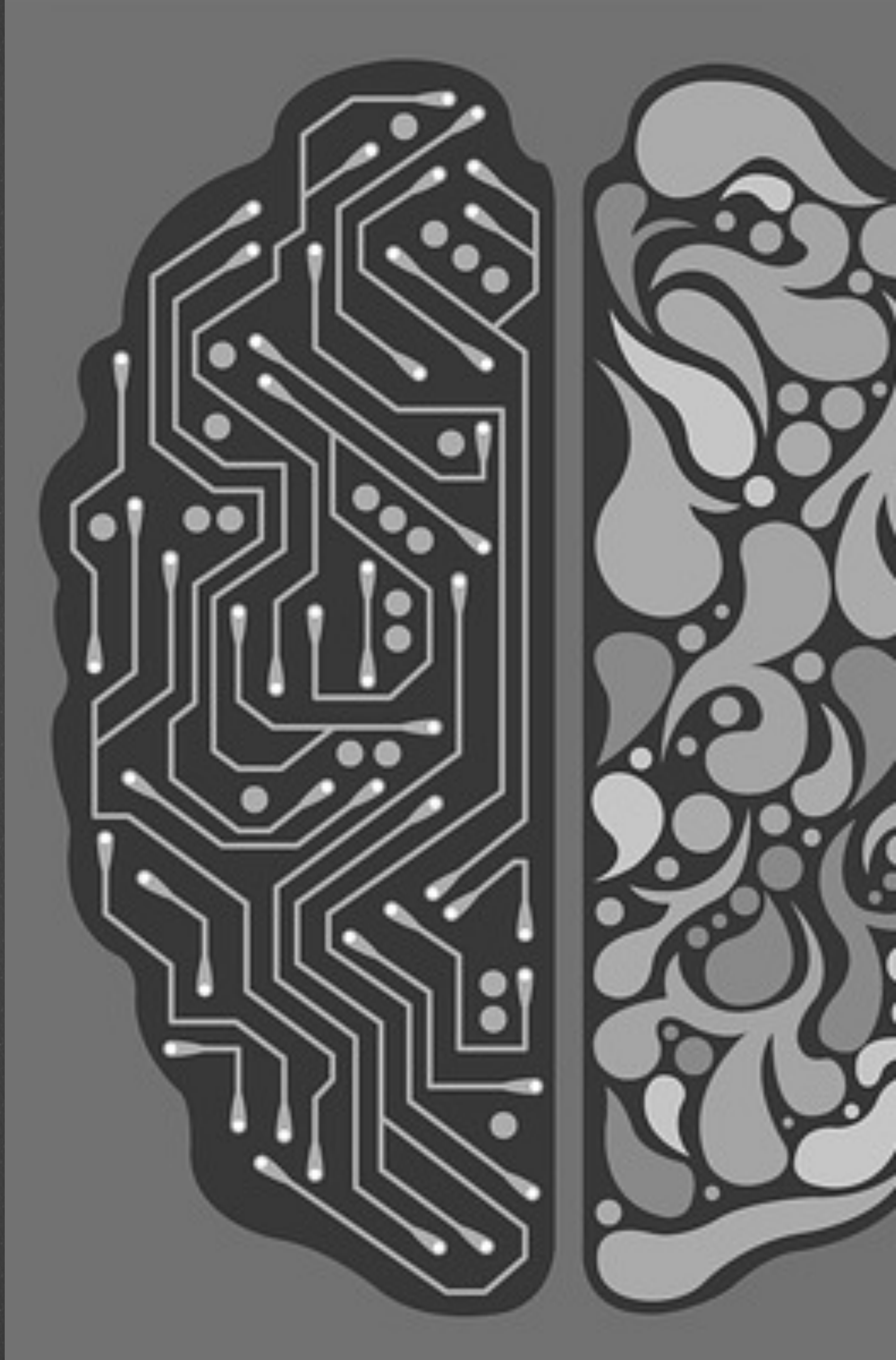
Data, data, data, …

02// Verifying MAS

Making sure the agent are working for us …

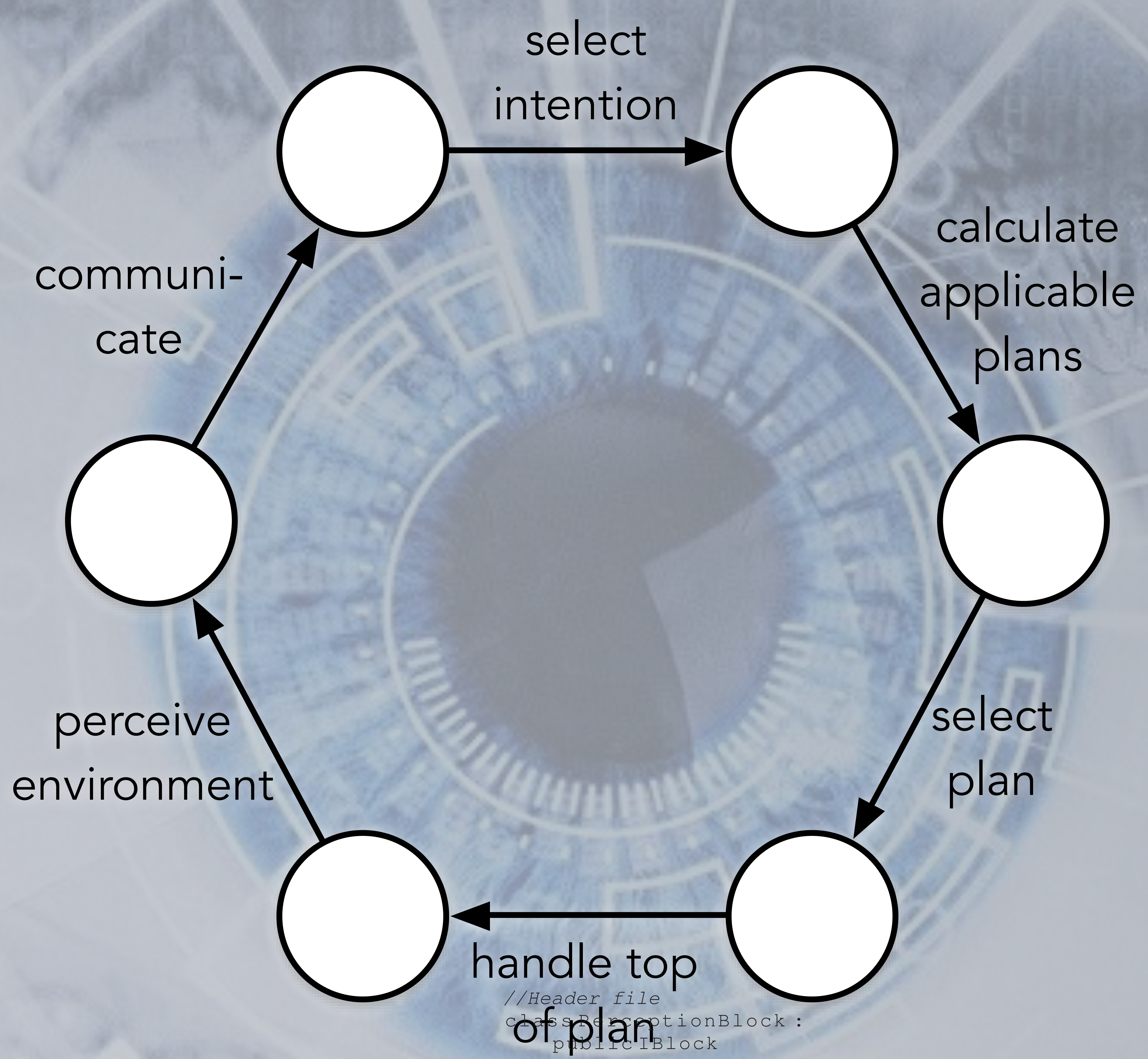
04// AI Futures

Academia’s responsibility in AI governance



Software Agents → Multi-Agent Systems

- Intelligent agents are characterised by:
  - autonomy
  - rationality
  - observation
- Agents can:
  - communicate
  - be proactive
- Agents can make a wrong decision.
  - Once realised, will attempt to find an alternative path.
- Why are we not seeing more agent technology?
  - What are the main challenges?
- **Dagstuhl Seminar, 2012:**
  - No tools
  - No OO support
  - No component-based approach
- Industry as an obstacle
  - No standards
  - No sharing of IP



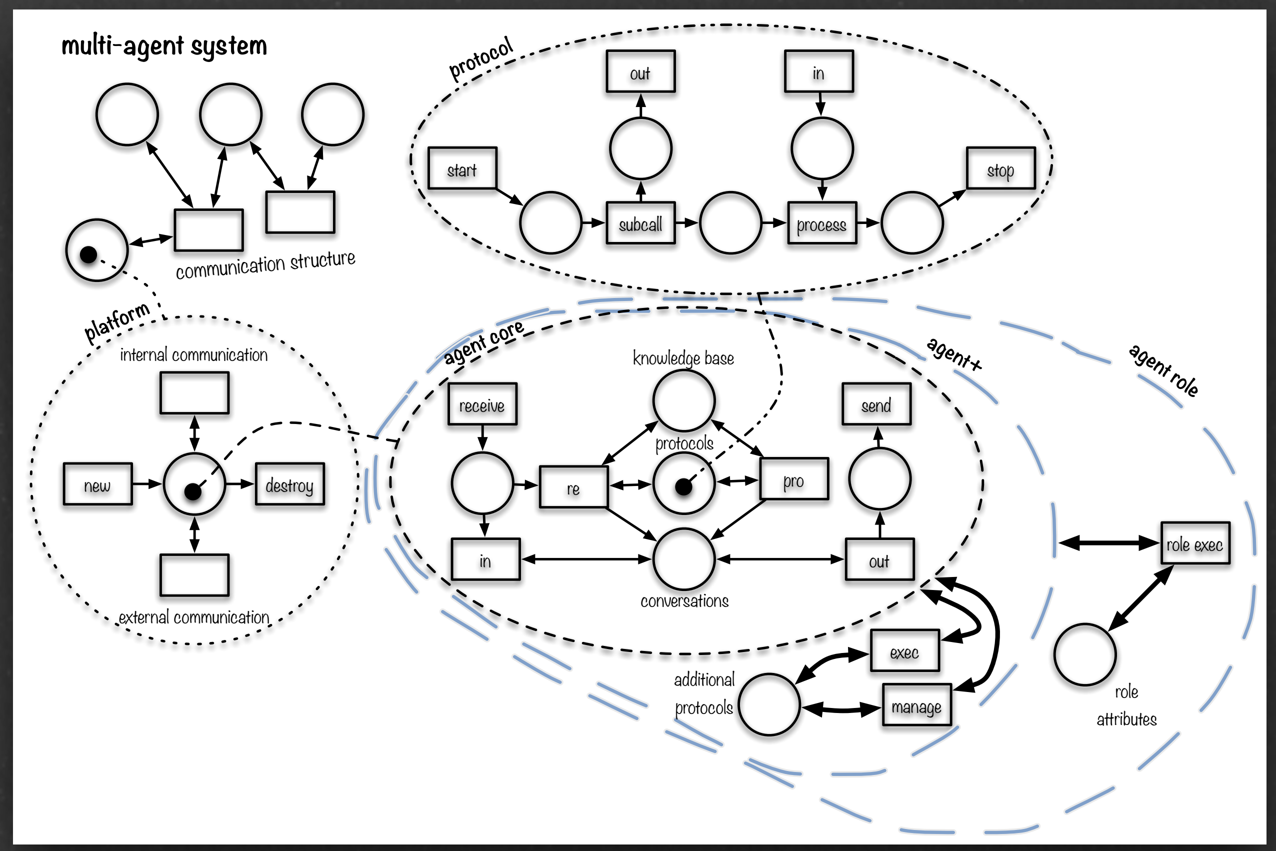
```
//Header file
class PerceptionBlock :
    public IBlock
{
public:
    PerceptionBlock(string name, vector<string> dependancies);
    PerceptionBlock(const PerceptionBlock& other);
    ˜PerceptionBlock();

    void Test(){}

    //Inherited function implementation
    void Setup(IBlock** depArr);
    void Run(EnvironmentState* envState);
};
```

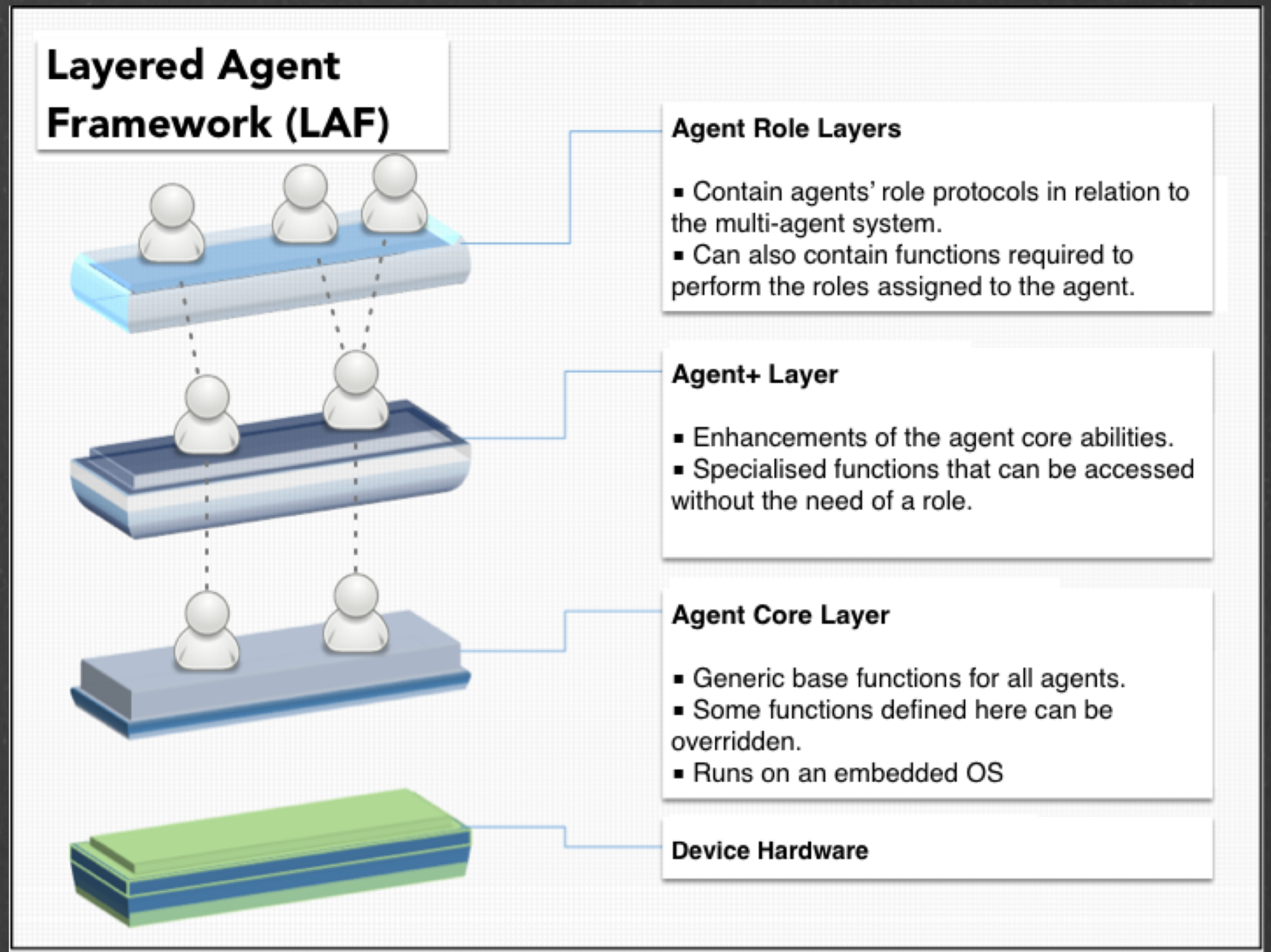
BDI

- **Beliefs**
  - Agent's knowledge
- **Desires**
  - Agent's goals
- **Intentions**
  - Plans that are being acted upon.



BDI Blocks

- Implemented in C++
- Use of database lookup
- Optimised database queries
- Tagging of beliefs with their origin(s)



Layered Agent Framework (LAF)

Agent Role Layers

- Contain agents' role protocols in relation to the multi-agent system.
- Can also contain functions required to perform the roles assigned to the agent.

Agent+ Layer

- Enhancements of the agent core abilities.
- Specialised functions that can be accessed without the need of a role.

Agent Core Layer

- Generic base functions for all agents.
- Some functions defined here can be overridden.
- Runs on an embedded OS

Device Hardware

Re-use & Debugging

- **Code Re-use**
  - Building blocks for BDI
  - Components for agent programming
  - OO Design Patterns

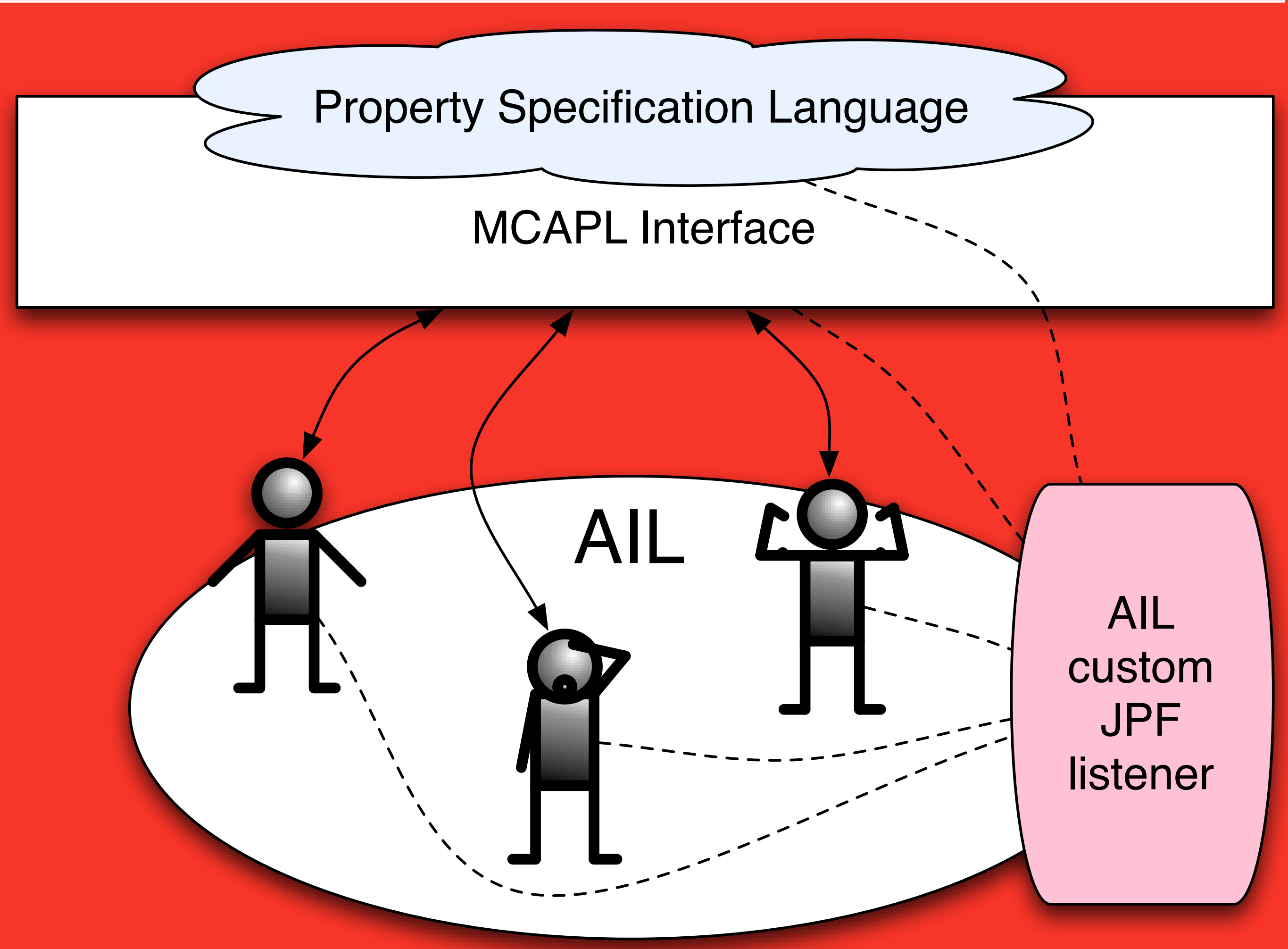
- **Debugging**
  - Extensive action/decision logs
  - Integrations with standard C++ debugging tools, e.g. in Visual Studio

- **Visualisation**
  - Drag-and-drop programming with connected blocks
  - Execution = Simulation
  - Inspection of (some) reasoning steps

- **Interfacing with existing Systems**

- One of the problems for MAS engineering is that academic approaches tend to use pure agent programming.
- Real-world applications require a model (design) that incorporates legacy components into an agent-equipped model.
- The move to an agent-assisted environment will not be instantaneous
- Industry 4.0

Verification



Agent Infrastructure Layer: intermediate language

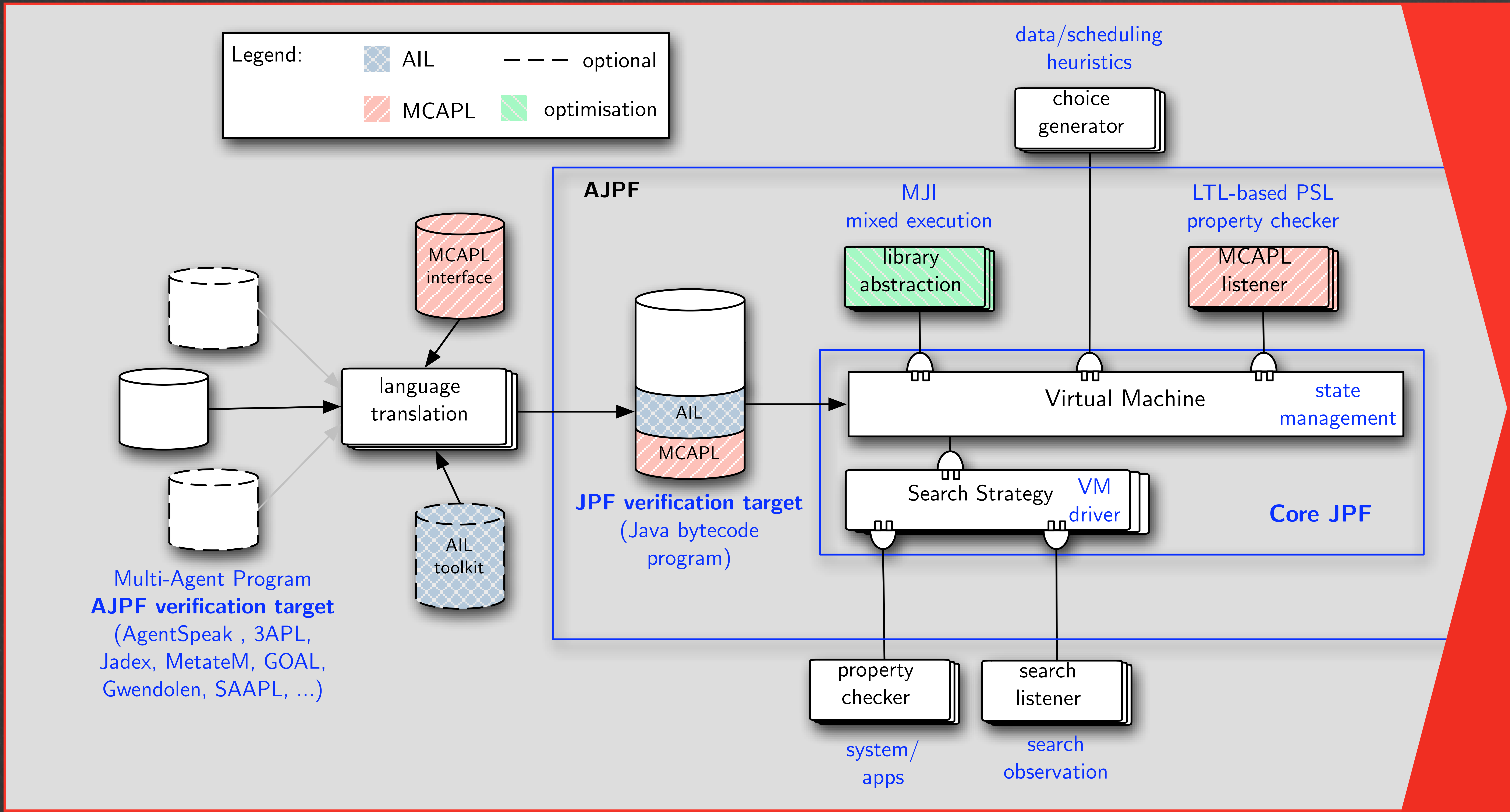
Expressing Properties

Property-specification language

Safe Componets

Use of agent libraries known to be safe and sound.

AJPF – Extending Java Pathfinder



Agent
Virtual
Machine

… a bit like a Java Virtual Machine

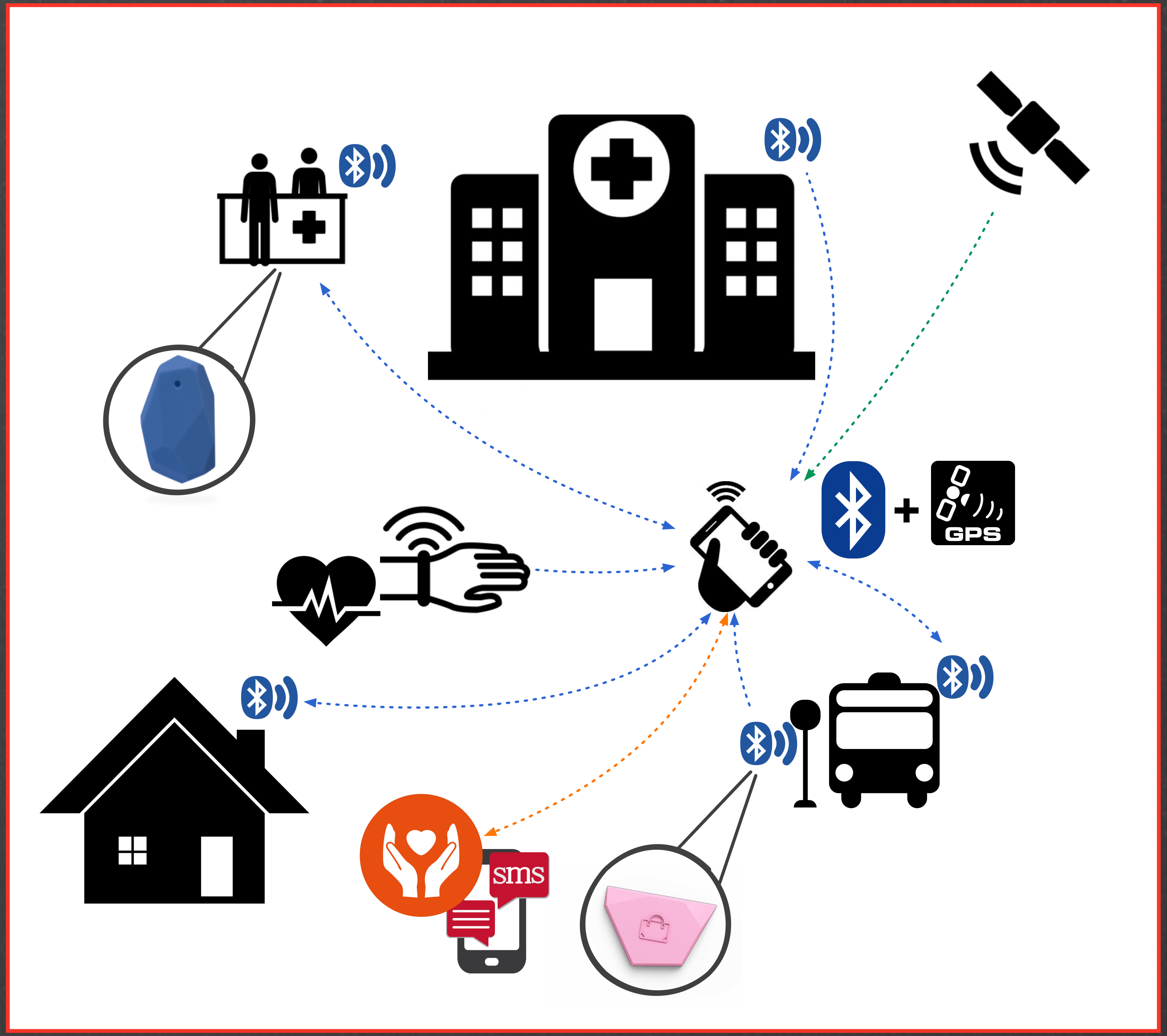
Trends & Challenges

- **Moving intelligence to ever smaller devices**
  - Mobility
  - (Dis-) Connectedness
  - Battery life

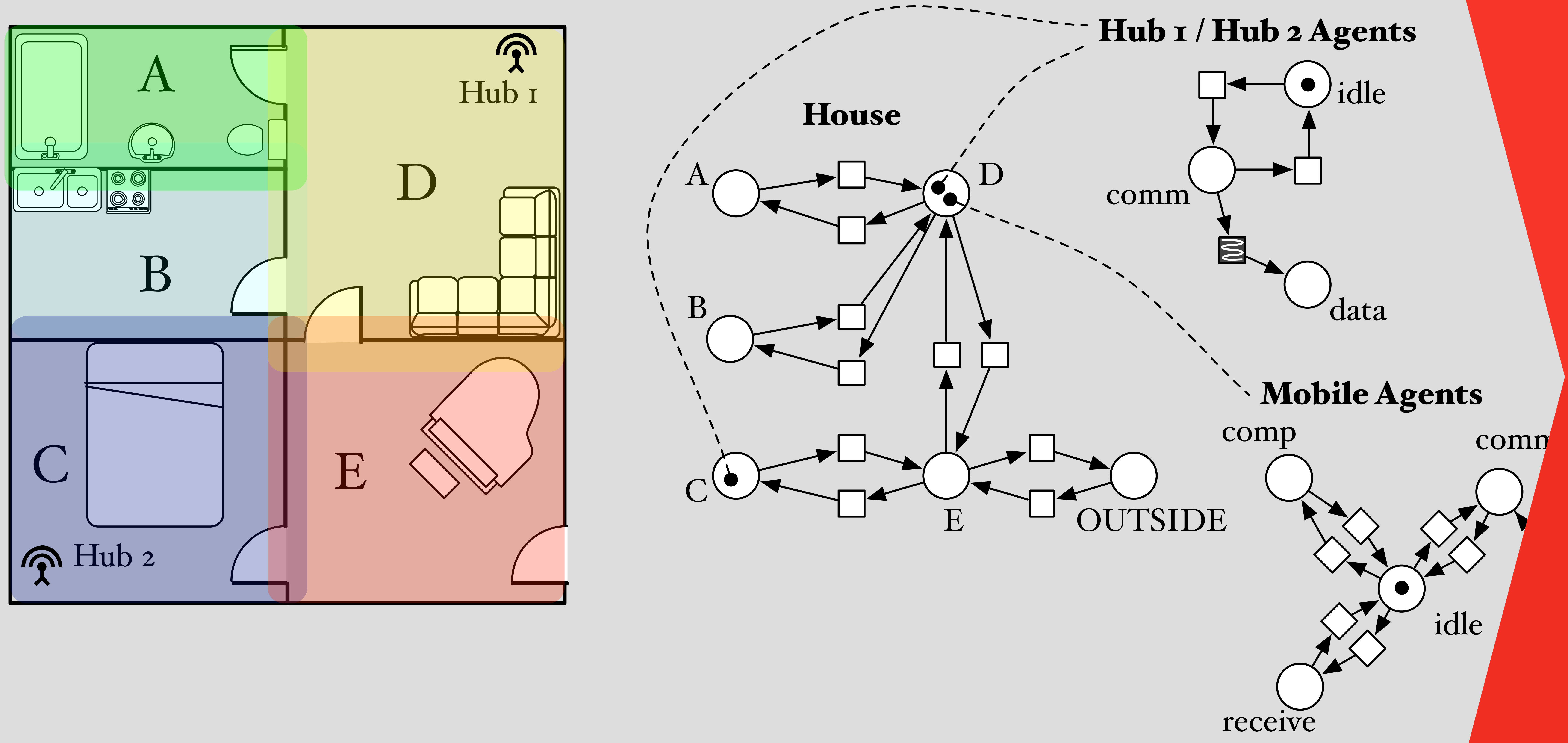
- **Ubiquity of sensors**
  - Accelerometers
  - Heart-rate sensors
  - GPS
  - Temperature

- **Using this trend for healthcare applicantions**

- Independence of Dementia patients



Smart Dementia Support



Wearables

Privacy, Ethics, and Risk Analysis embedded in a mobile device reacting to its enviroom=nments

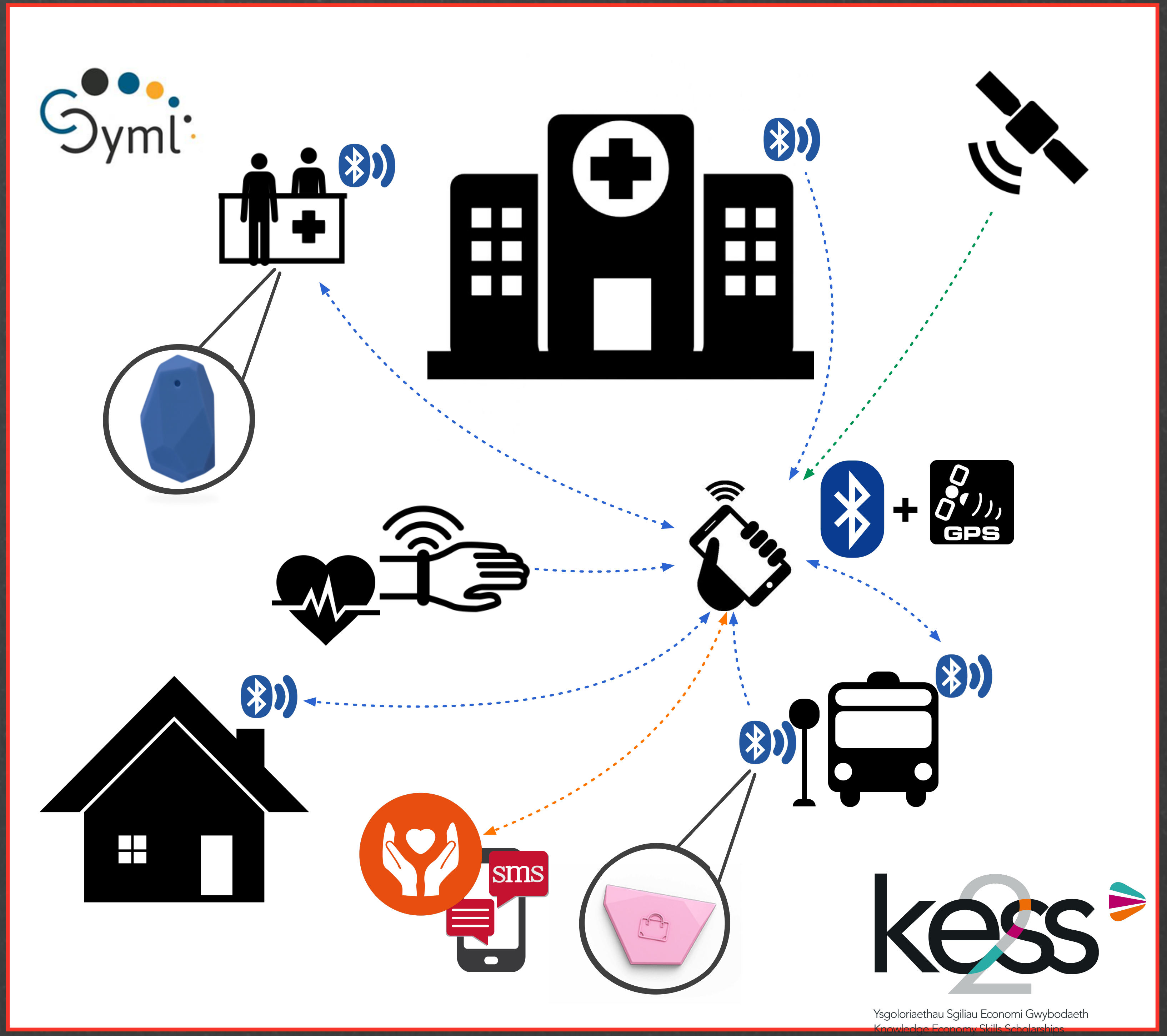
Patient Monitoring

- **Machine Learning**
  - Locations
  - Sleep
  - Behaviour

- **Agents**
  - Activity sensing
  - Environmental factors
  - Dynamic risk assessment
  - Dynamic ethics assessment

- **System is completely unobtrusive**

- Carers can be alerted if thresholds warrant disclosure of data


Syml
GPS
sms
kess 2
Ysgoloriaethau Sgiliau Economi Gwybodaeth
Knowledge Economy Skills Scholarships

AI – Sustaining Current Success

What needs to be done to make these successes last?

Building on availability of data and powerful analysis tools, we are now faced with privacy concerns and tighter legal requirements (GDPR).

FUTURE
CARE
CAPITAL

Taking concerns seriously is paramount to continued success.

aippg appg
ARTIFICIAL INTELLIGENCE

BIG
INNOVATION
CENTRE
APPG AI SECRETARIAT

Personalised AI

AI with humanity in mind

Dennis R. Mortensen
CEO and founder, x.ai

> AI isn’t very good at jobs that require creativity, empathy, critical thinking, leadership, artistic expression, and a whole host of other qualities we traditionally think of as “human.”

Combat Fears

AI isn’t as smart as you may think

What needs to be done to make AI the successes many already think it is?

Raising public awareness about what can and what cannot be achieved with AI (currently and in the future).

Transparency and taking concerns seriously is paramount to continued success.

Aaron Levie
CEO, Box

> “AI can seem dystopian because it’s easier to describe existing jobs disappearing than to imagine industries that never existed appearing.”

Give Reassurance

Humans are smarter than you think

What will we do to tackle the problems that automation will bring upon us?

Create new jobs, think up new industries.

Facilitate creativity to deal with the “job crisis”.

Offer Visions

AI =
Augmented Intelligence

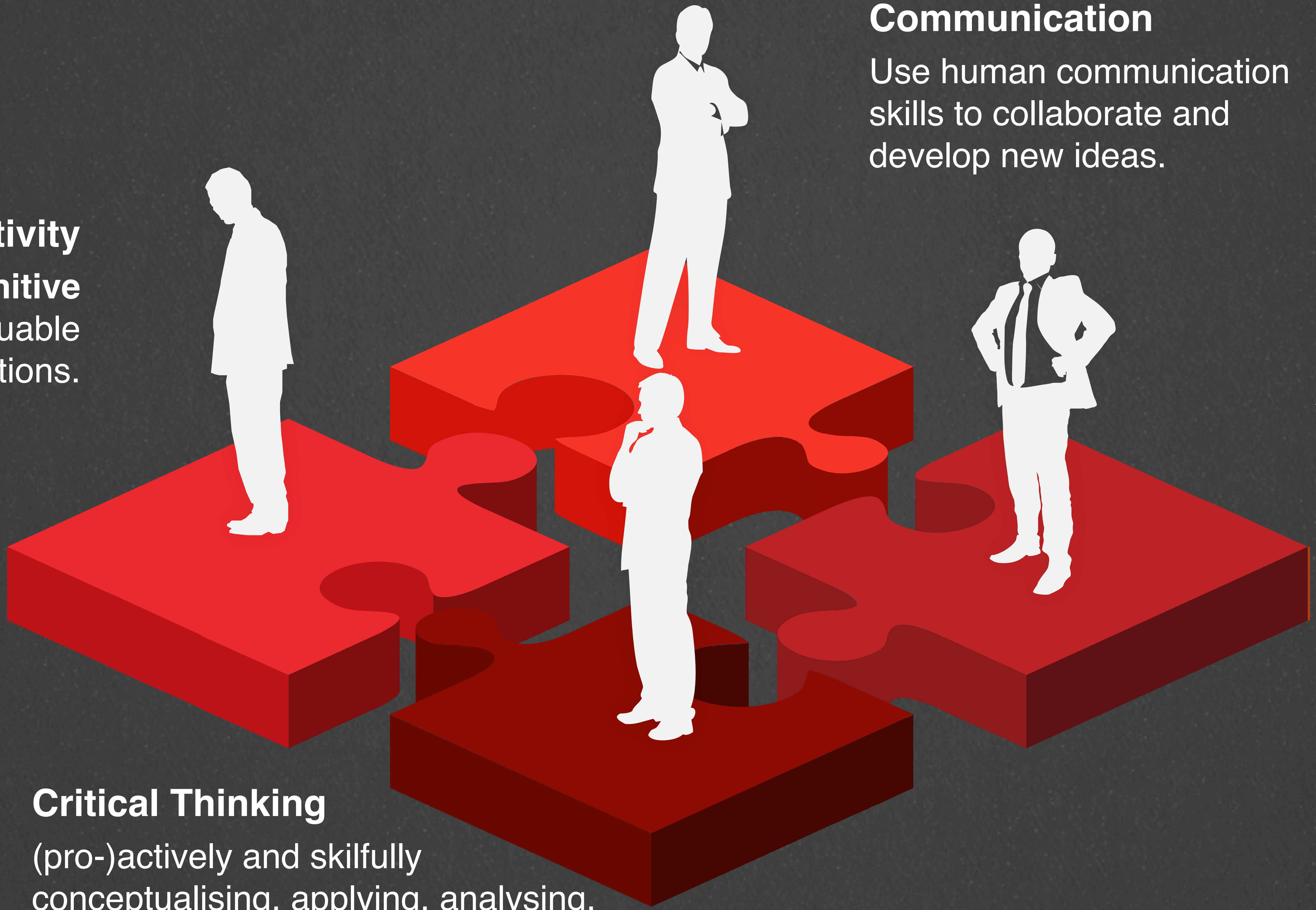
Make AI work for and with human intelligence.

Design systems to support rather than replace human intelligence.

Create human jobs to support AI and address accountability.

Reduce mundane tasks and improve customer & workforce satisfaction without reducing the workforce.

Human Qualities



Communication

Use human communication skills to collaborate and develop new ideas.

Creativity

Together with **cognitive flexibility**, create valuable innovations.

Emotional Intelligence

Ability to join intelligence, empathy and emotions to enhance thought and understanding of interpersonal dynamics.

Critical Thinking

(pro-)actively and skilfully conceptualising, applying, analysing, synthesising, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action

Federated AI

Distributed data acquisition

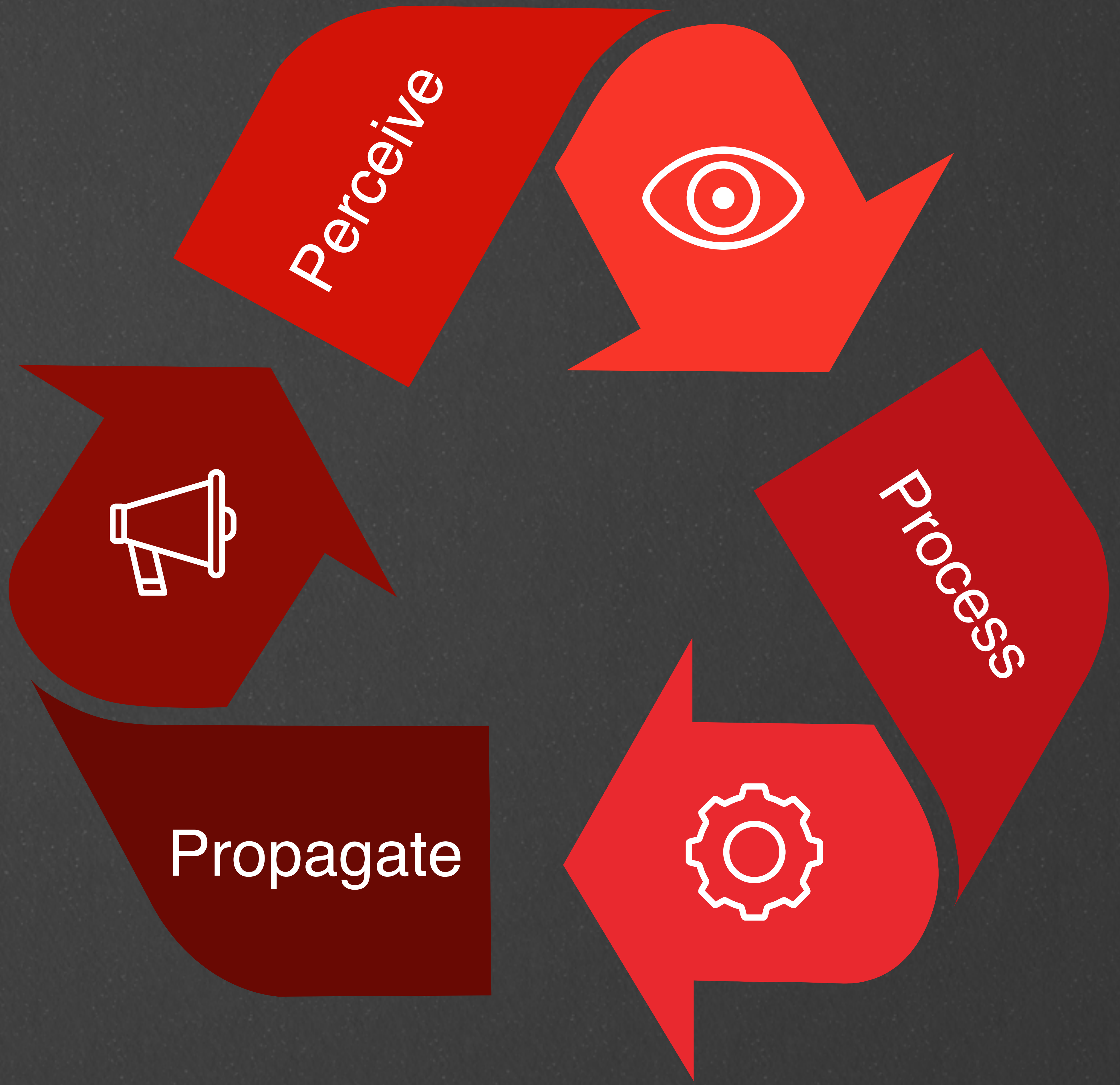
- E.g., on mobile phones.
- Millions of decentralised nodes.
- Compression to deal with bandwidth problems.

Local processing - Federated learning

- Locally compute updates.
- Iterative model averaging.

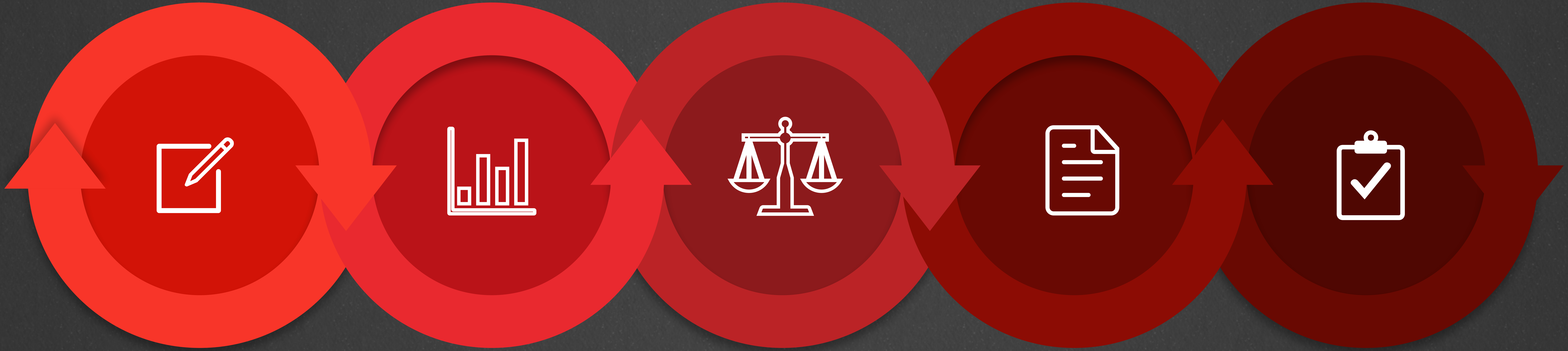
Centralised consolidation & propagation

- Communicate updates.
- Securely aggregate information.
- Apply deep learning.



Dynamic Ethical Reasoning

Re-consider ethics at ‘runtime’ in the current context, e.g., in health applications and policing.



Thresholds
Definition of ethical thresholds

Weightings
Dynamic weighting of environmental factors

Law
Consideration of legal requirements and restrictions

Contract
Agreement on permissible use of data

Decision
Final decisions checked against the contract

-Eliezer Yudkowsky-

> “By far the greatest danger of Artificial Intelligence is that people conclude too early that they understand it.”

TRUST

Privacy

All parties need to be able to rely on basic rights of privacy

Consent

All parties need to have an agreement in place stating mutual consent about the way data is processed

Tracability

Decision making process needs to be transparent and traceable

If we don’t understand it, how can we trust it?

AI needs to work for and with humanity. This can only be achieved on a basis of trust.

Approved Contextual Compliance for AI

RESPONSIBILITY → VOLATILITY → OBLIGATION → ETHICS

Responsibility

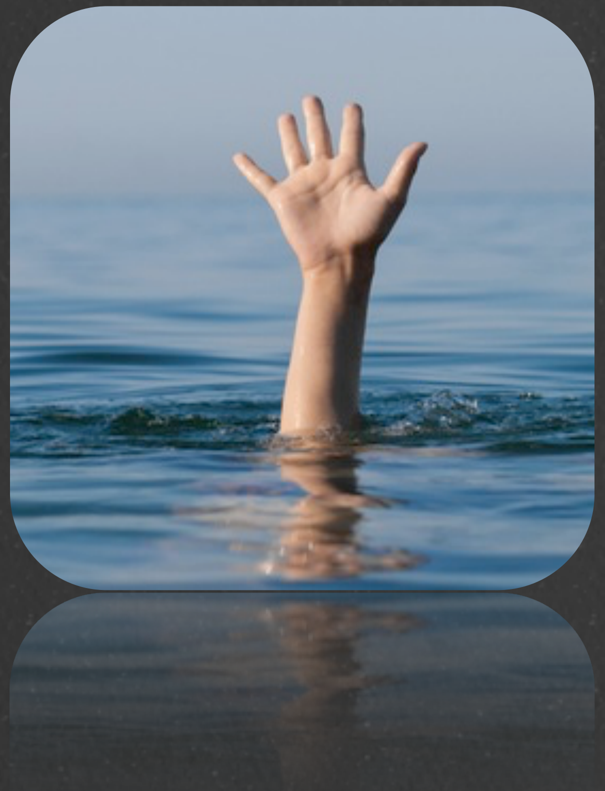
Sensitivity of data can increase by adding public data that would not be regarded as sensitive on its own

→ responsible handling and combination of data

Volatility

Capability of adapting to changing environments without re-programming

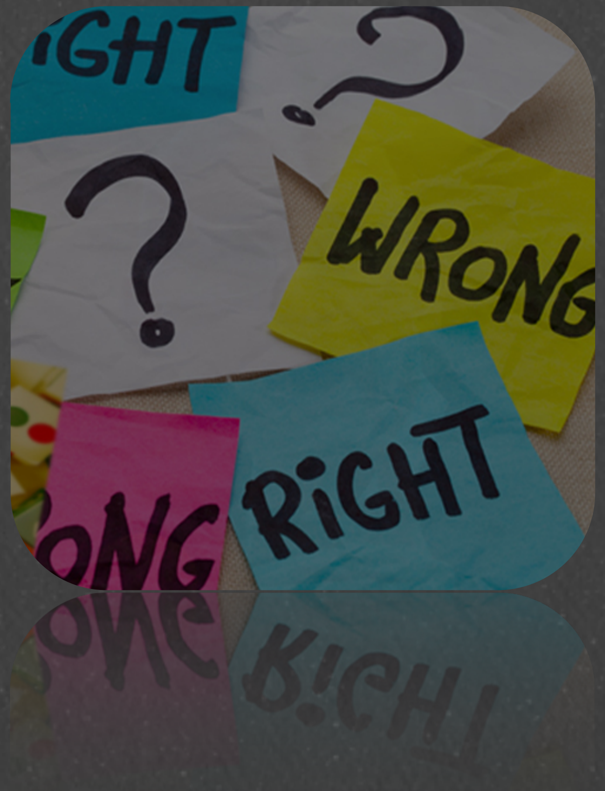
→ proactive rational agency



Obligation

Ability and obligation to act on perception of current context

→ contractual actions



Ethics

Ethics need to be engineered into AI systems. Relying on our responsible use of AI is not enough.

→ 'book of ethics' incorporated into AI

Work on AI Standards

Regulated AI?
Yes please!

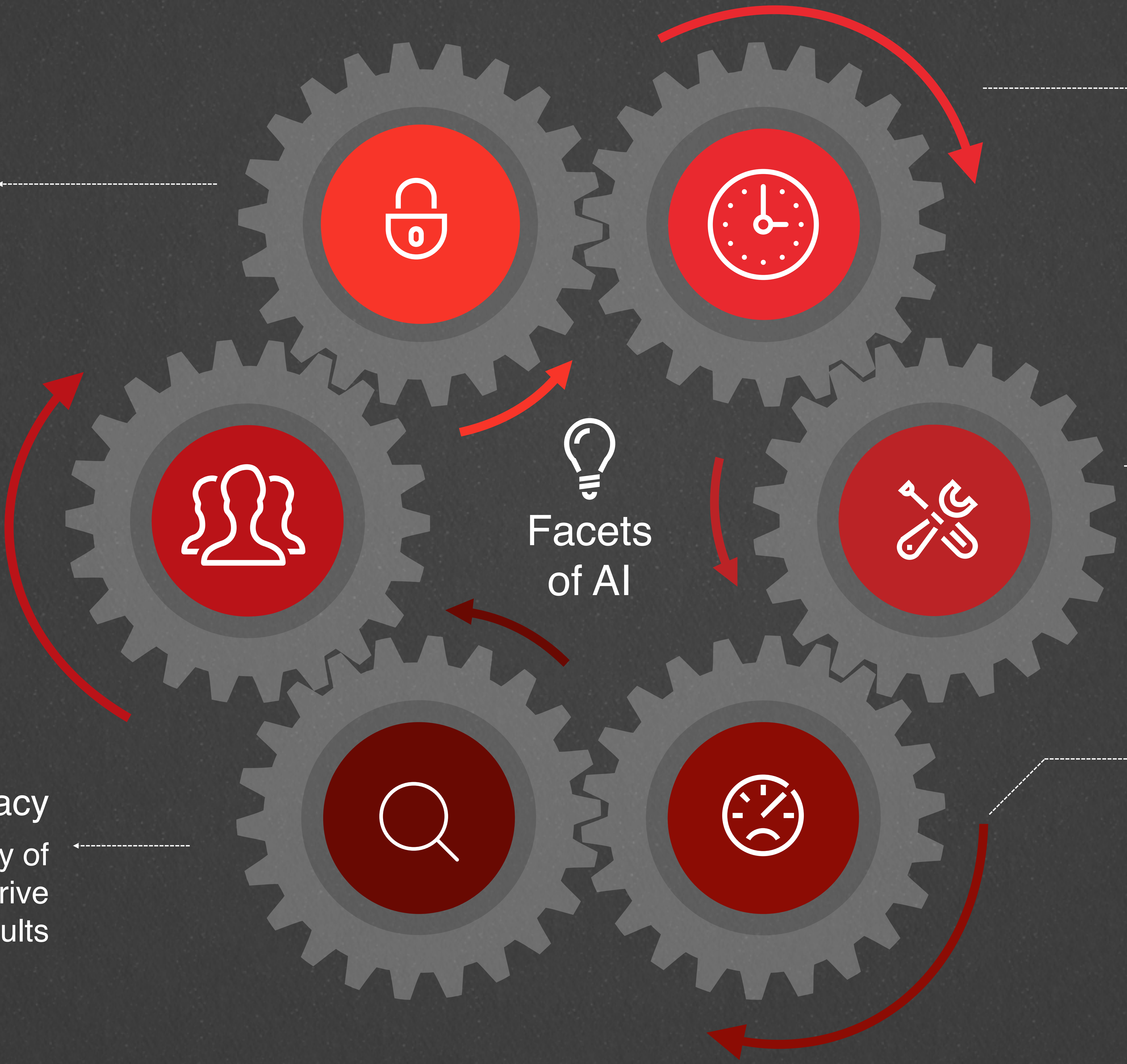
Are the legal and ethical problems specific or general, national or international?

Clearly, some are international and need to be solved by international standards and regulations.

Other industries have thrived from the introduction of standards, so why not AI as well?

Address questions of data governance and accountability.

The Future of Ubiquitous AI



Security

Protection against hacking and data corruption.

Availability

Services available at any time and from a mobile device

Acceptance

Users view results as appropriate and ethically approve of

Effectiveness

Methods reliably deal with any given situation and provide an answer

Accuracy

Correctness and adequacy of data and methods to derive results

Efficiency

Good use of resources and timely results.

The Future of Ethical AI

Possible?

A lot is possible with today's AI technologies, but should we use them just because we can?

Legal?

National and International laws regulate data collection, data storage and data processing.

Ethical?

Ethical arguments need to be taken into account. These are dynamic and the context can override a previous argument.

Cost-effective?

Financially viable and computational feasible.

Responsible?

Data accuracy and adequacy of methodology as well as recognition of privacy are paramount.

Accepted?

Consent must be sought. Transparency is key to user acceptance.

APPG AI Santa Challenge

Any Questions?

“Wise old **Santa uses AI** to find the right present for everyone. His **AI uses contextual information**, such as the culture, religion, and the meteorological season (to name just a few aspects), to find the most suitable Christmas or year-end presents. This could be a new *BBQ for Susan in Sydney*, the *latest 'Oseibo’ for Toshihiro in Tokyo*, a *sled for Henry in Hampshire*, or some *ingredients for 'nyama choma' the traditional Christmas meal for Kwamboka from Kenya*. To ensure the AI produces accurate results, the **Elves work hard** at checking that the underlying datasets have as **little bias** as possible. They have also spent a lot of time and effort to **make the AI's decisions transparent**, so Santa can trust their recommendations to avoid being embarrassed at the disappointment when the present wasn’t what the recipient had wished for.”