

Human-Like AI

**From logic to argumentation: reasoning with
imperfect knowledge in the era of AGI**

Dave Raggett, W3C/ERCIM



Why is Human-Like AI Interesting?

Human-like AI will be hugely disruptive to web search, personal privacy, and ecosystems of services*

- ❑ Personal agents that safeguard your privacy, help you with your health, financial affairs, education and so forth
- ❑ Agents to influence you, agents to defend against that
- ❑ Agents communicating with other agents to find and provide services in open decentralised ecosystems
- ❑ Agents on the Web, in the Metaverse, as robots and embedded in other devices, including cars
- ❑ The *intelligent* Web of Things!

* As well as human society in the large, requiring political change to ensure we all benefit

AI is advancing rapidly

Any sufficiently advanced technology is indistinguishable from magic, Arthur C. Clarke

- ❑ Some examples of recent successes based upon neural networks and deep learning
 - [ChatGPT](#) – akin to a conversational version of Wikipedia and lots more
 - [Minerva](#) – undergraduate science problems
 - [Stable Diffusion](#) – image generation
 - [Google Translate](#) – natural language translation
- ❑ Self-supervised learning from huge datasets using expensive hardware, and big budgets with a big carbon footprint
 - Smaller, more agile systems are coming, optimised for the edge-cloud continuum
 - Adapting large models to new applications using transfer learning is much less costly
- ❑ Lots of opportunities for research on the **Intelligent Web of Things** as agents that can communicate, learn, and reason like we do, supporting human-machine collaboration to boost productivity
- ❑ Applying insights from **cognitive sciences**
- ❑ **Ethical considerations**, see [FastAI's introduction to data ethics](#)
- ❑ *“Breakthrough work in deep learning absolutely does not require access to vast resources, elite teams or advanced math. There is lots of work still to be done that requires just a bit of common sense, creativity, and tenacity.”*, Jeremy Howard and Sylvain Gugger (FastAI)

Symbolic AI is falling behind relative to Neural Networks

But has a crucial ongoing role in respect to interoperability

- ❑ **Symbolic AI** focuses on formal semantics and logical proof
- ❑ In practice, though, most reasoning is hard coded in applications, and lacks flexibility
- ❑ Greater flexibility in conjunction with sub-symbolic metadata, e.g. PKN*
- ❑ Explicit knowledge provides for effective transparency
- ❑ *Hard to scale* up due to dependency on expensive hand-crafted knowledge
- ❑ **Neural networks** are good for plausible reasoning with everyday knowledge that is uncertain, imprecise, incomplete and inconsistent
- ❑ Argumentation for and against a premise, just like in courtrooms and everyday discussions
- ❑ Challenges around transparency and provenance in respect to explainability
- ❑ Neural networks are *easy to scale* up using larger datasets, automating knowledge engineering, and trainable for specific tasks

* Plausible Knowledge Notation, a level above RDF that embraces fuzzy logic and qualitative reasoning

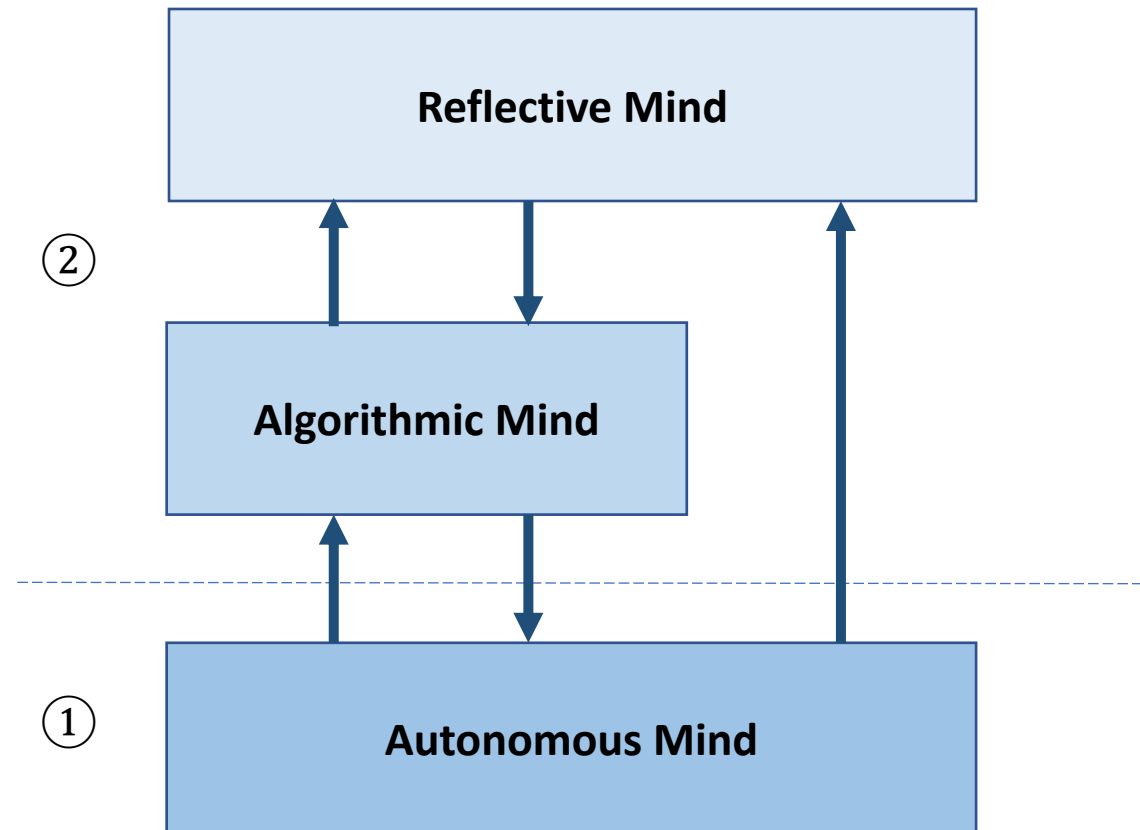
Neurosymbolic AI seeks to combine strengths of both disciplines

Keith Stanovich's Tripartite Model of Mind

Type 2 processing is slow, deliberative, and open to introspection, e.g. mental arithmetic. It is formed by chaining Type 1 processes using working memory.

Type 1 processing is fast, automatic, and opaque, e.g. recognising a cat in a photograph or a traffic sign when driving a car.

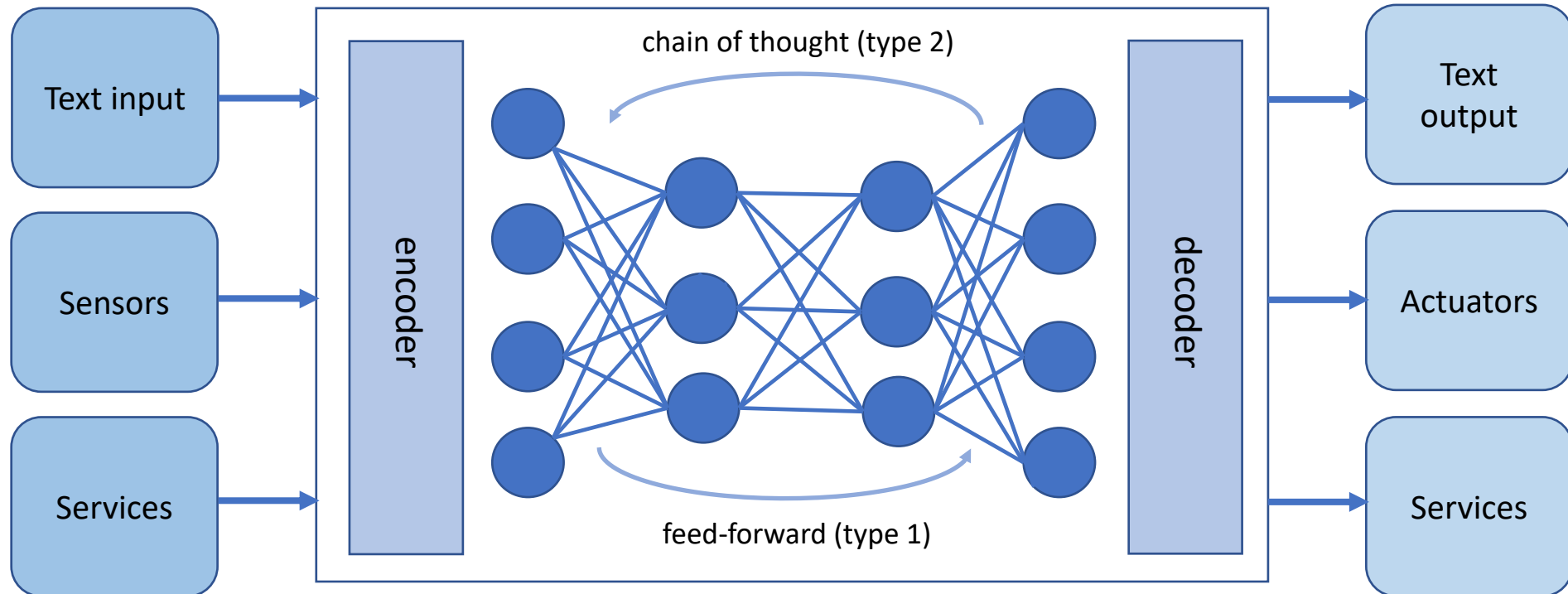
Sometimes referred to as System 1 & 2



See, e.g. “Dual-Process Theories of Higher Cognition: Advancing the Debate”, Evans and Stanovich (2013), along with “Thinking Fast and Slow”, Daniel Kahneman (2011)

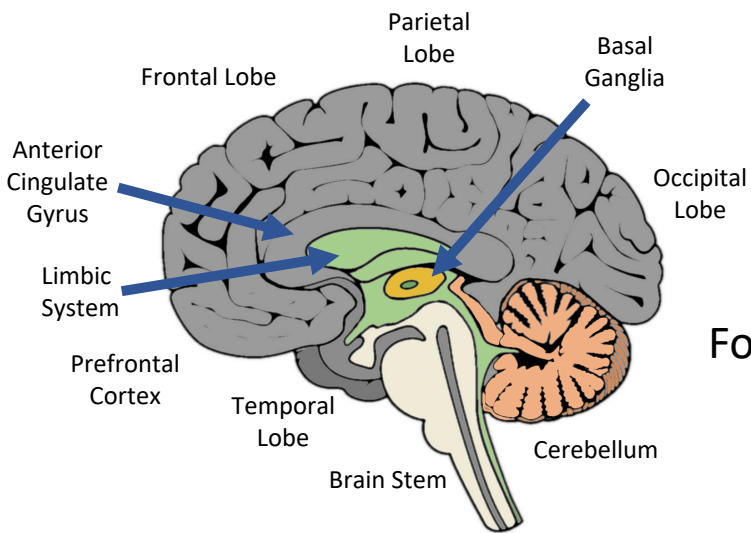
Architecture for Neurosymbolic Cognitive Agents

Combining intelligence with back-end IT systems



Services include cognitive databases and reasoners using, e.g. PKN, along with scripts and tools to generate tables, charts and other graphics. Actions are delegated to external real-time control loops.

The diagram depicts a high level neural architecture for cognitive agents, based upon reinforcement learning with human feedback, as used for today's large language models. In theory, it could use comparatively smaller models as they would each be trained for a specific application area. Reasoning is based upon chain of thought processing, along with asynchronous access to external services. The network in the diagram is iconic and not intended as an accurate representation - something too hard to draw in a simple diagram.



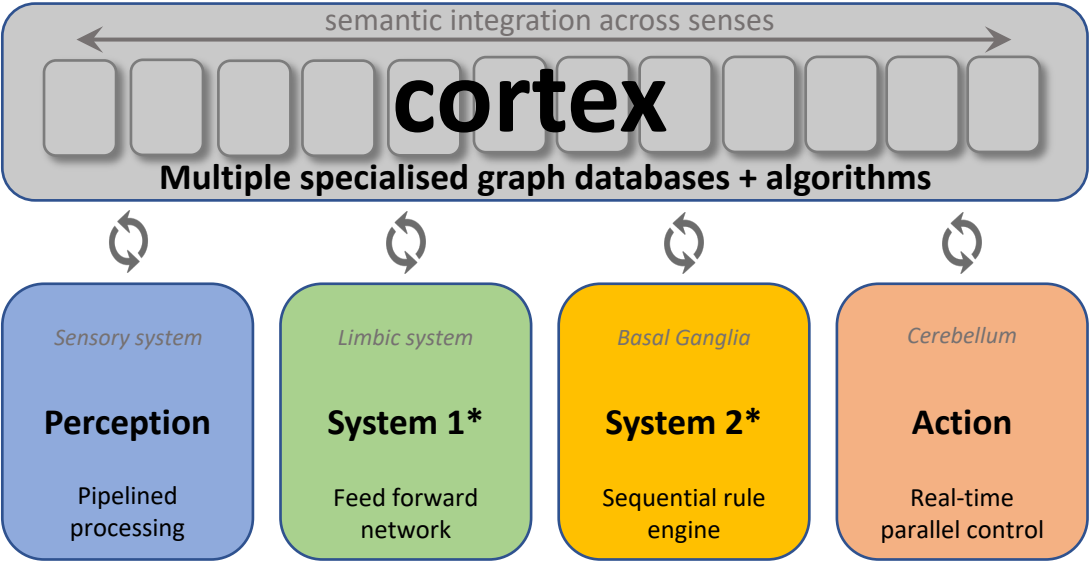
The temporal lobe acts as a hub for integration across the senses

Cognitive Architecture for artificial minds

For both Symbolic and Neural Network implementations



Multiple cognitive circuits loosely equivalent to shared blackboard



Cortex supports memory and parallel computation. Recall is stochastic, reflecting which memories have been found to be useful in past experience. Spreading activation and activation decay mimics human memory with semantic priming, the forgetting curve and spacing effect. Hub and spoke model for semantic integration across senses.

Perception interprets sensory data and places the resulting models into the cortex. Cognitive rules can set the context for perception, and direct attention as needed. Events are signalled by queuing chunks to cognitive buffers to trigger rules describing the appropriate behaviour. A prioritised first-in first-out queue can be used to avoid missing closely spaced events.

System 1 covers intuitive/emotional thought, cognitive control and prioritising what's important. The limbic system provides rapid assessment of past, present and imagined situations. Emotions are perceived as positive or negative, and associated with passive or active responses, involving actual and perceived threats, goal-directed drives and soothing/nurturing behaviours.

System 2 is slower and more deliberate thought, involving sequential execution of rules to carry out particular tasks, including the means to invoke graph algorithms in the cortex, and to invoke operations involving other cognitive systems. Thought can be expressed at many different levels of abstraction, and is subject to control through metacognition, emotional drives, internal and external threats.

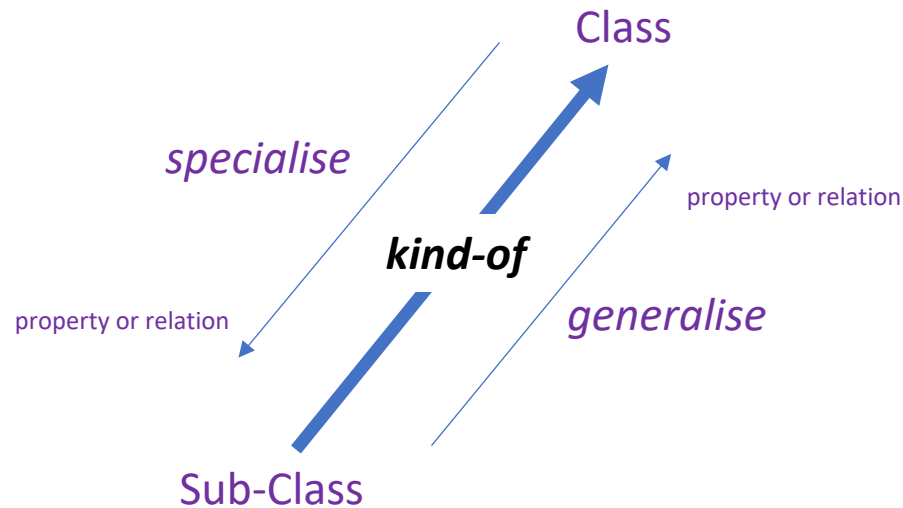
Action is about carrying out actions initiated under conscious control, leaving the mind free to work on other things. An example is playing a musical instrument where muscle memory is needed to control your finger placements as thinking explicitly about each finger would be far too slow. The cerebellum provides real-time coordination of muscle activation guided by perception. It further supports imagining performing an action without carrying it out.

* Equivalent terms to Type 1 and 2 processing

From Logic to Argumentation

Guidelines for effective arguments have been worked upon by many philosophers since the days of Ancient Greece, e.g. Walton's argumentation schemes

- Inferring likely properties and relations across other relations



- Expected certainty influenced by qualitative metadata
 - e.g. *typicality*, *similarity*, *strength* that represent prior knowledge as gut feelings in lieu of detailed statistics

- Forward and backward inferences using implications

weather is raining *implies* weather is cloudy

- If it is raining then it is cloudy
- If it is cloudy it may be rainy

- Inferences based upon analogies

- matching structural relationships

- Imprecise concepts

- fuzzy terms, e.g. cold, warm and hot
- fuzzy modifiers, e.g. *very* old
- fuzzy quantifiers, e.g. *few* and *many*

- Multiple lines of argument for and against the premise in question

Plausible Reasoning (ChatGPT)

Scalar ranges with fuzzy terms:

If the room is cold, is it warm?

No, if a room is cold, it is not warm. Cold refers to a low temperature, while warm refers to a high temperature. If a room is cold, it means the temperature inside the room is lower than the desired or comfortable temperature.

If the room is hot, is it warm?

Yes, if a room is hot, it can be considered warm. Hot refers to a high temperature and warm refers to a comfortable or desirable temperature which is usually high. So, a room that is hot, is also warm.

Reasoning by similarity:

Belgium is good for growing daffodils. England has a similar climate to Belgium. Is England good for growing daffodils?

It is likely that England has a similar climate to Belgium and therefore may also be good for growing daffodils. However, there are many other factors that can affect the suitability of a location for growing a particular plant, such as soil quality, available sunlight, and local pests and diseases. It is best to consult with a local expert or conduct research on local growing conditions to determine if daffodils can be successfully grown in England.

Large language models do pretty well at plausible reasoning, and furthermore, can justify their reasoning

Plausible Reasoning (ChatGPT)

Love triangle:

John loves Mary, Mary loves Peter, but not John, is John jealous of Peter?

It is possible that John may feel jealous of Peter if he knows that Mary loves Peter and not him. Jealousy is a complex emotion and can be influenced by various factors such as past experiences, self-esteem, and individual personality. Without more information, it is difficult to say for certain whether John is feeling jealous of Peter.

Reverse implications:

If it is cloudy, is it rainy?

It is not necessarily rainy if it is cloudy. Cloudy weather can indicate a variety of conditions, such as the presence of clouds but no precipitation, or the possibility of rain or other types of precipitation in the near future. It depends on the weather conditions and the specific area you are referring to.

Simple analogical puzzles:

petals:flowers:wheels:?

petals:flowers::wheels:vehicles

bees:hive::bears:?

bees:hive::bears:den

Why?

The analogy is comparing the relationship between bees and their hive, to the relationship between bears and their den. Just as bees live and work together in a hive, bears also live and raise their young in a den. The analogy is showing that both bees and bears have a specific place where they live and operate within their environment.

Large language models do much better than systems based upon symbolic knowledge

Plausible Reasoning (ChatGPT)

Basic Math:

What is x given that $18 = 2/x^2$?

We can find x by rearranging and taking the square root of both sides:

$$18 = 2/x^2$$

$$x^2 = 2/18$$

$$x^2 = 1/9$$

$$x = \pm\sqrt{1/9}$$

$$x = \pm 1/3$$

So x can be either $1/3$ or $-1/3$.

This example demonstrates chain of thought, i.e. type 2 processing.

Human-like errors with unverified assumptions, *oops!*

The party has 15 visitors, 8 of them doctors and the rest of them actors. Dave Raggett and George Clooney join the party. How many actors are there now?

Before Dave Raggett and George Clooney joined the party, there were $15 - 8 = 7$ visitors who were actors. After they joined, there were $7 + 2 = 9$ actors at the party.

Is Dave Raggett an actor?

I'm not aware of Dave Raggett being an actor. According to my training data, Dave Raggett is a computer scientist and software engineer who has worked on the development of the World Wide Web and has been involved in standardization efforts related to the web.

So why did you assume he was an actor?

My apologies, I made an error. I assumed Dave Raggett was an actor based on the context of the question and the presence of George Clooney, who is a well-known actor. I should have verified that information before answering.

How would you match all this using symbolic reasoning and the Semantic Web? Does it make sense to even try?

Known Issues for Deep Learning

- ❑ Generative AI is amazing but makes stupid mistakes, e.g. factual and logical errors, inconsistencies, limited reasoning, toxicity, fluent and convincing hallucinations, ...
- ❑ Generative AI has no common sense and can't plan answers
 - Statistical prediction without reflective reasoning
 - Human reflect on and revise their texts as they work on them
- ❑ Humans are easily fooled by the fluency of LLMs and project our own fears onto them
 - Need to inform people they are chatting with an AI, and to educate them about its limitations
- ❑ Deep learning suffers from catastrophic forgetting, precluding continuous learning
 - Prompt engineering from external sources as a weak work around

Research Challenges

- ❑ Distilling LLMs to better suit modest sized systems, and as a basis for transfer learning for tasks of interest
- ❑ Faster and smarter learning with smaller datasets – mimicking human learning
 - Self-guided learning (Type 1 and 2 cognition)
 - Combining generalisation and memorisation
- ❑ Taming catastrophic interference as key to enabling AGI
 - Transfer learning & neural network generalisability across tasks
- ❑ Introspection, metacognition and reflective reasoning
 - Involving Type 2 processing and working memory
- ❑ Earning trust through citing provenance, avoiding careless mistakes and offensive responses
 - Good habits, reflective thinking, fact checking
 - Self-critique on biased or offensive responses
- ❑ Integrating episodic memory and continual learning: instruction, observation and direct experience
- ❑ Human languages with limited training resources
 - Transfer learning from languages with many speakers
- ❑ Distributed hive minds for faster learning across agents with shared access to cognitive databases
 - Knowledge gained by individual agents is immediately accessible to all agents
 - The hive is strengthened by a diversity of agent personalities and skills, increasing resilience to changing environments
- ❑ Agent-Client confidentiality
 - Key to ensuring privacy
 - Policy based data sharing
 - Application to privacy-centric ecosystems of services
 - Emotional intelligence in respect to human interaction
 - Learning and applying behavioural norms
- ❑ Adapting human societies to ensure everyone benefits – regulations and tax incentives
 - [EU AI Act](#) imposes a ban on unacceptable risk, and legal requirements on high risk applications

There are many opportunities for improving on today's LLMs for Human-like AGI

Reflective Cognition

Reflective cognition will enable faster learning and the ability to provide better explanations

❑ Reflection covers

- Choosing to reflect on a question rather than give the first response that comes to mind
- Providing a commentary as you work on a problem
- Thinking, comparing, deciding, including what are the best ways to approach a given problem
- Thinking about learning, including identifying problems, proposing and evaluating ideas
- Theory of mind for self and others, including beliefs, desires and intentions

❑ Skill acquisition

- Starts with deliberate reasoning that is slow and error prone
- Further practice yields fast, automatic performance
- Skills build one upon another
- Children can only describe (i.e. reflect on) what they are doing after acquiring a skill

❑ What is needed for reflective cognition in neural networks?

- Episodic memory as a basis for self-awareness and learning from the past

Behavioural Norms

- ❑ To be an effective partner for human-machine collaboration, cognitive agents will need to follow behavioural norms for social interaction
- ❑ Simplest examples include mimicry, e.g. smiling back when someone smiles at you
- ❑ Gricean maxims as norms for cooperative dialogues
- ❑ We readily detect when someone violates accepted norms, and may apply sanctions
- ❑ Implicit norms are those that are automatic
- ❑ Explicit norms are those that are written down
- ❑ How can cognitive agents acquire norms and detect violations?

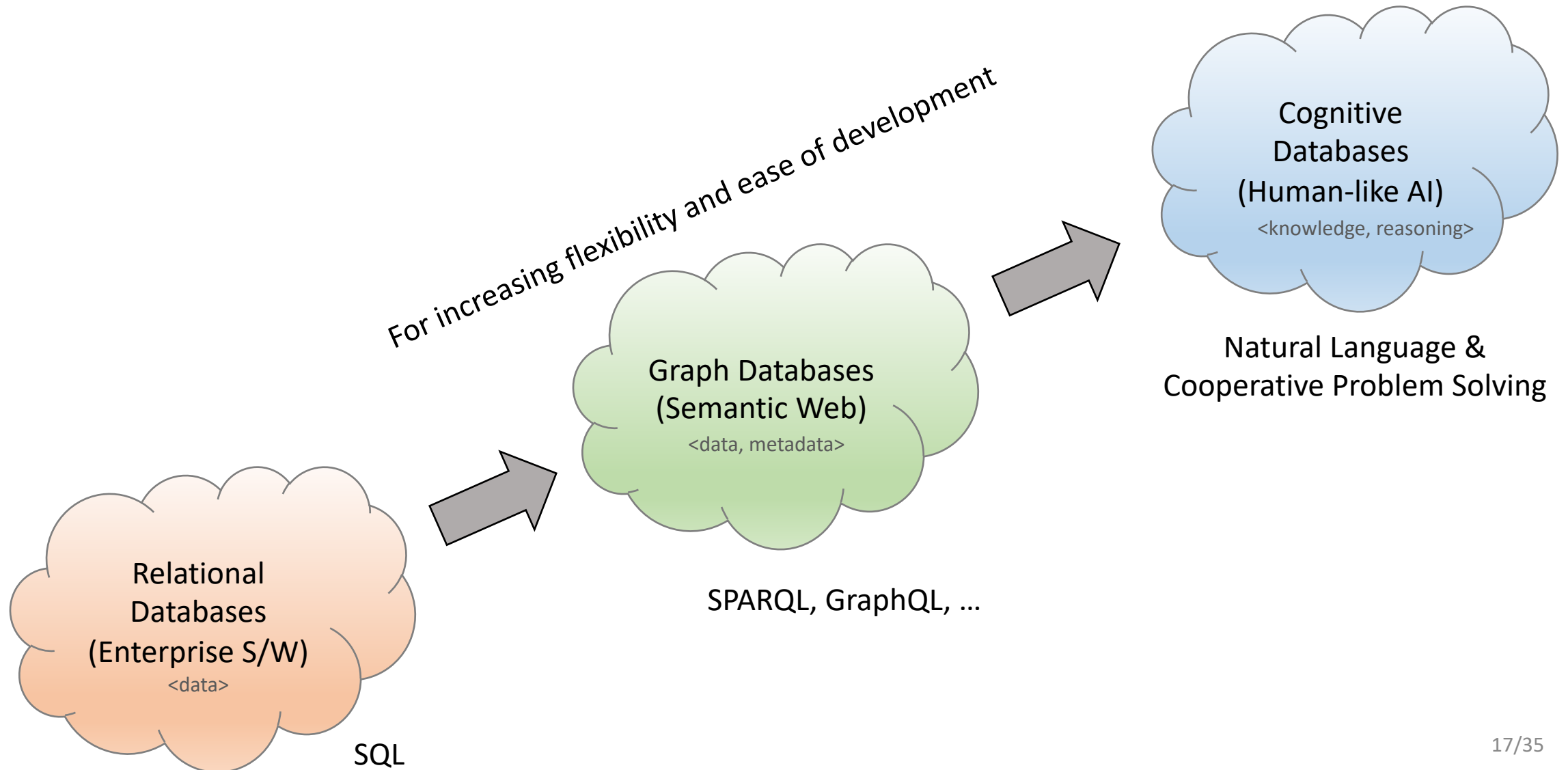
Related to theory of mind, causal reasoning, explaining and predicting behaviours

Effective Argumentation

The long list of argumentation schemes proposed by Walton emphasise the need for cognitive agents (and people) to learn how to reason effectively.

- ❑ How to teach large language models and symbolic systems to reason effectively?
- ❑ How to measure the veracity and strength of arguments?
- ❑ How to tailor arguments to a specific audience?
- ❑ How to assess how well a system or person argues?
 - How to detect unintended biases and intentional but hidden (dog-whistle) tropes?
- ❑ Evolutionary race between systems designed to influence people and systems designed to defend against that – impact on social networks and politics
- ❑ A big question is how to develop systems capable of learning how to reason when given suitable examples together with annotations describing their strengths and weaknesses
 - Need to generalise beyond the given examples
- ❑ Could we pre-train symbolic language models akin to how neural network models are pre-trained using masked word prediction?
- ❑ But if neural networks are a particularly effective technology for this, then we would be best to focus on them, right?
- ❑ There is a lot to be gained by studying best practices for teaching human students.

Evolution in ICT Systems



Going Beyond the Limitations of Today's Symbolic AI

- ❑ Traditional logic draws inferences from a set of facts and rules where things are either true or false
- ❑ Inconsistencies are not allowed!
- ❑ But the real world is imperfect with knowledge that is uncertain, imprecise, incomplete and inconsistent
- ❑ RDF is W3C's framework for representing symbolic knowledge
- ❑ But it is based upon traditional logic and suffers the same limitations
- ❑ The Plausible Knowledge Notation (PKN) is designed to support imperfect knowledge

How can we improve on RDF?

Some examples of PKN from [web-based demo](#)

For symbolic knowledge, PKN offers more flexible semantics, and an easier to use notation, relative to RDF/turtle & JSON-LD

- *PKN statements: properties, relations and implications*
- *Qualitative statement metadata and scopes*
- *Imprecise concepts with fuzzy scalars, fuzzy logic, fuzzy modifiers and fuzzy quantifiers*
- *Analogical reasoning*

climate **of** Belgium **includes** temperate

guilt **of** accused **excludes** guilty

roses **kind-of** temperate-flowers

circuit **analogous-to** plumbing

flow **increases-with** pressure **for** plumbing

current **increases-with** voltage **for** circuit

weather **of** ?place **includes** rainy

implies weather **of** ?place **includes** cloudy (**strength** high, **inverse** low)

leaf **part-of** tree

up **opposite-to** down

Mary **younger-than** Jenny

younger-than **equivalent-to** less-than **for** age

range **of** age **is** infant, child, adult **for** person

age **of** infant **is** birth, 4 **for** person

John **loves** chess

subject **of** loves **includes** person

object **of** loves **includes** hobby (**strength** medium)

dog:puppy::cat:kitten

bird:flock::fish:?

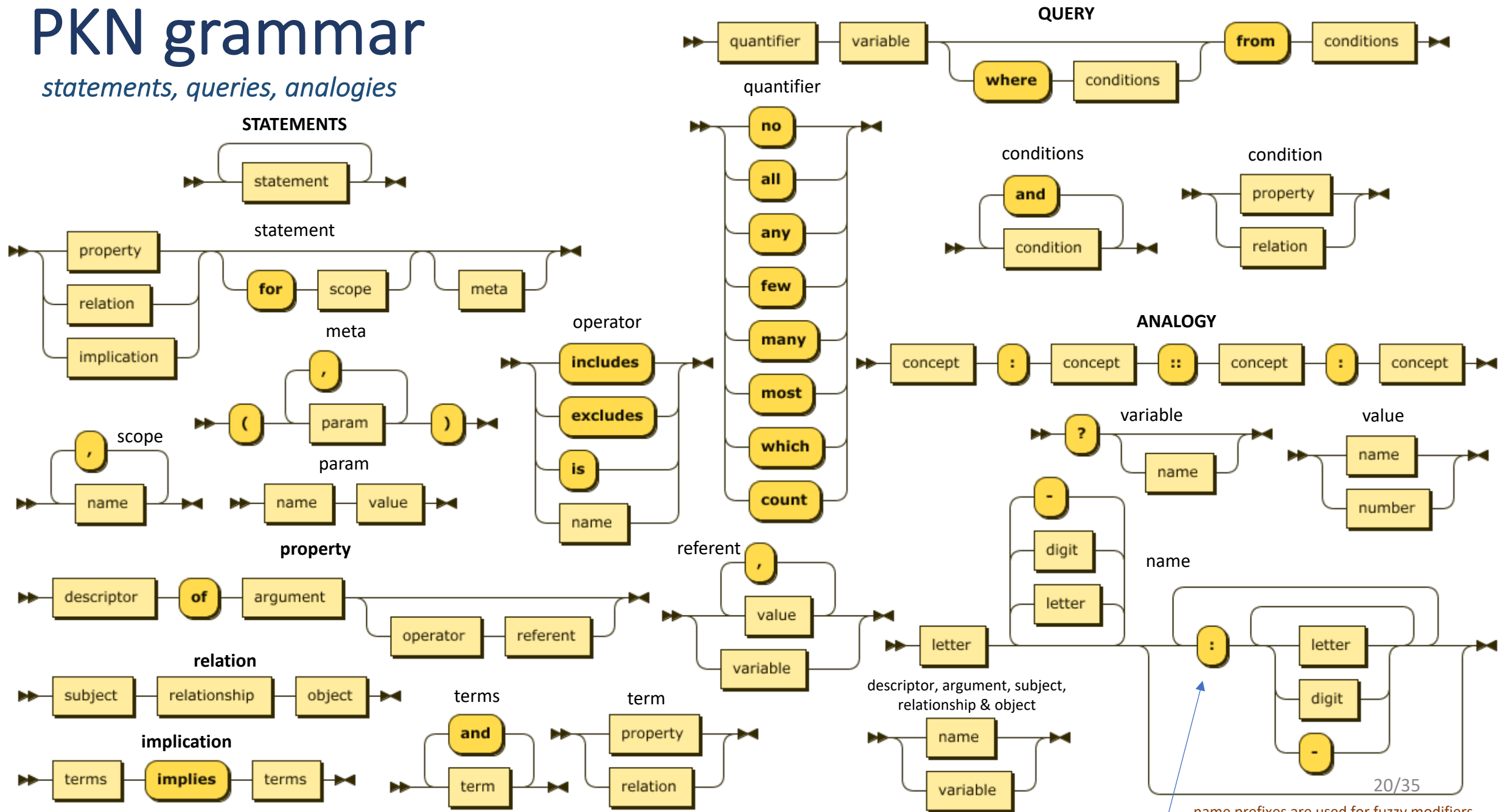
which ?x **where** ?x **is-a** person **and** age **of** ?x **is** very:old

count ?x **where** age **of** ?x **greater-than** 20 **from** ?x **is-a** person

few ?x **where** color **of** ?x **includes** yellow **from** ?x **kind-of** rose

PKN grammar

statements, queries, analogies



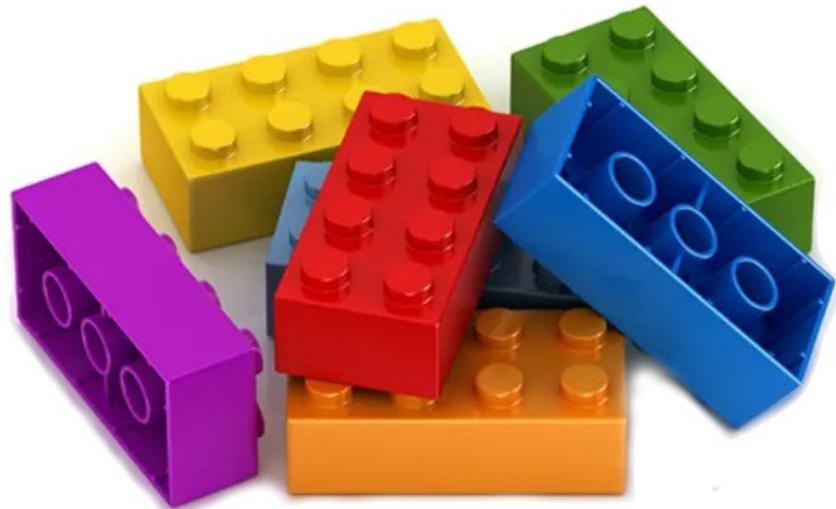
How does PKN relate to RDF and Neural Networks?

- ❑ W3C's Resource Description Framework (RDF) is based upon triples, i.e. labelled directed graph edges
 - <subject, predicate, object>
- ❑ PKN statements can be modelled as a collection of triples
- ❑ Consider: flowers of England includes roses (certainty high)
- ❑ This can be mapped to an RDF blank node as the subject for 5 triples*
- ❑ PKN uses comma separated lists for collections – something that is harder to express in RDF
- ❑ PKN can be considered as a member of a family of notations that provide richer semantics compared to RDF
- ❑ PKN can also be mapped to cognitive chunks – sets of name/value pairs
 - e.g. as used by CMU's cognitive architecture: [ACT-R](#) ‡
- ❑ Large language models internally represent chunks as vectors of activation values in spaces with high dimensions
- ❑ This explains how artificial neural networks can manipulate semantic graphs in their working memory

* using context files analogous to JSON-LD to define the mapping from names to URIs

‡ See also chunks and rules from the [W3C Cognitive AI CG](#)

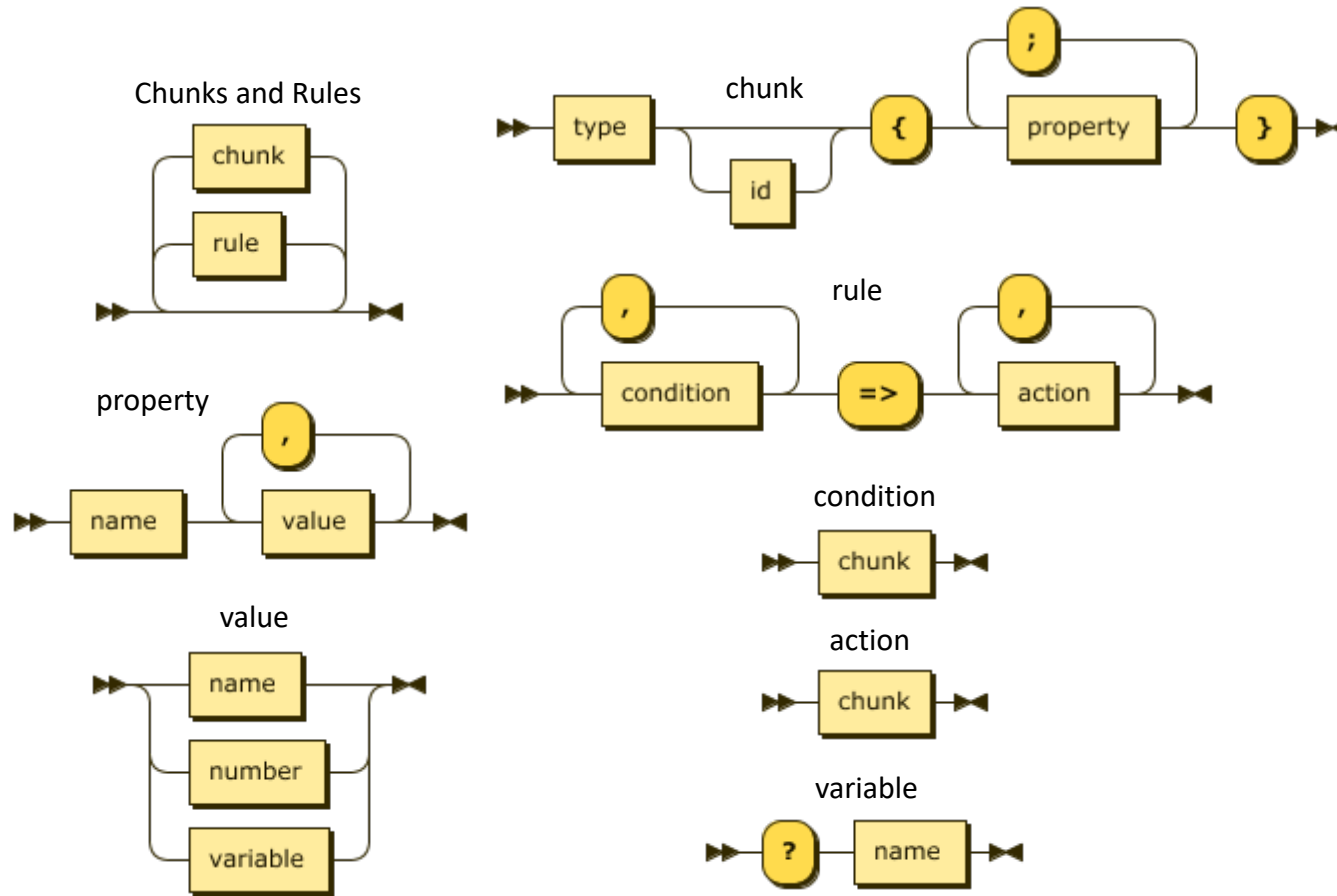
Chunks and Rules



- ❑ RDF is sometimes thought to be hard to work with, due in part to the focus on triples (graph edges) rather than higher level concepts
- ❑ Property Graphs are dubbed as a little easier and have seen increasing adoption by industry
- ❑ Chunks and Rules are easier still, and builds upon decades of research in the cognitive sciences
- ❑ Simple syntax for representing and working with properties, i.e. sets of name/value pairs, where names and values can reference other chunks
- ❑ Playing a complementary role to PKN

Chunks and Rules

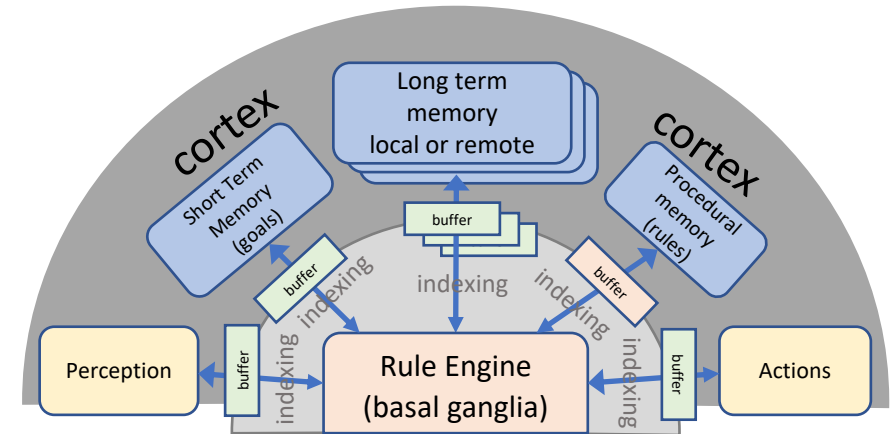
web-based demos for smart homes and factories



names beginning with “@” are reserved, e.g. @do for actions

See [W3C Cognitive AI Community Group](#)

Cognition – Sequential Rule Engine



Cognitive Buffers hold single chunks
Analogy with HTTP client-server model

- Inspired by John Anderson’s ACT-R
- Mimics characteristics of human cognition and memory, including spreading activation and the forgetting curve*
- Rule conditions and actions specify which cognitive module buffer they apply to
- Variables are scoped to the rule they appear in
- Actions either directly update the buffer or invoke operations on the buffer’s cortical module, which asynchronously updates the buffer
- Predefined suite of cortical operations
- Reasoning decoupled from real-time control over external actions, e.g. a robot arm

* We forget so we can focus on what’s important 13/35

Robot Control – Smart Factory

<https://www.w3.org/Data/demos/chunks/robot/>

- **Cognitive AI demo that runs in a web page**
- Live simulation of bottling plant with robot, conveyor belts, filling and capping stations
- Real-time control by a cognitive agent

```
# add bottle when belt1 has space and wait afresh
```

```
space {thing belt1} =>
```

```
    action {@do addBottle; thing belt1},
```

```
    space {@do wait; thing belt1; space 30}
```

```
# stop belt1 when it is full and move arm
```

```
full {thing belt1} =>
```

```
    action {@do stop; thing belt1},
```

```
    action {@do move; x -120; y -75; angle -180; gap 40; step 1}
```

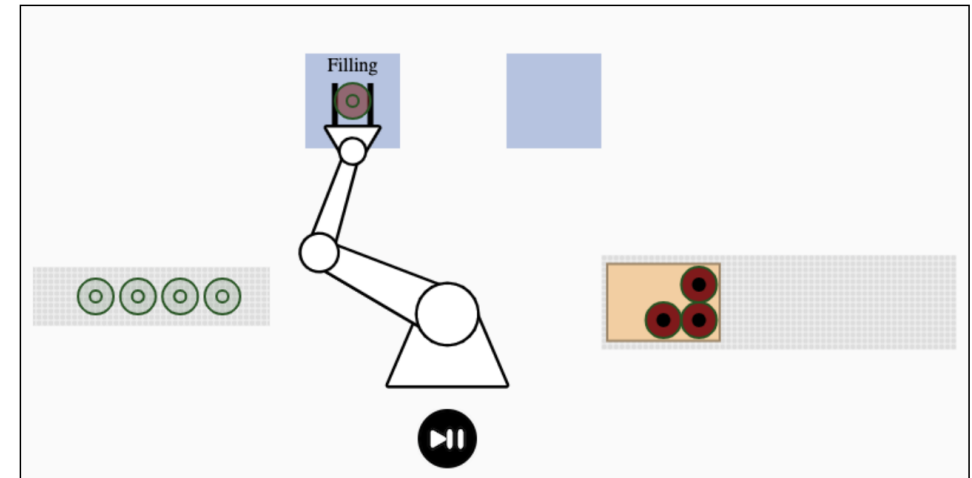
```
# move robot arm into position to grasp empty bottle
```

```
after {step 1} => robot {@do move; x -170; y -75; angle -180; gap 30; step 2}
```

```
# grasp bottle and move it to the filling station
```

```
after {step 2} => goal {@do clear}, robot {@do grasp},
```

```
    robot {@do move; x -80; y -240; angle -90; gap 30; step 3}
```



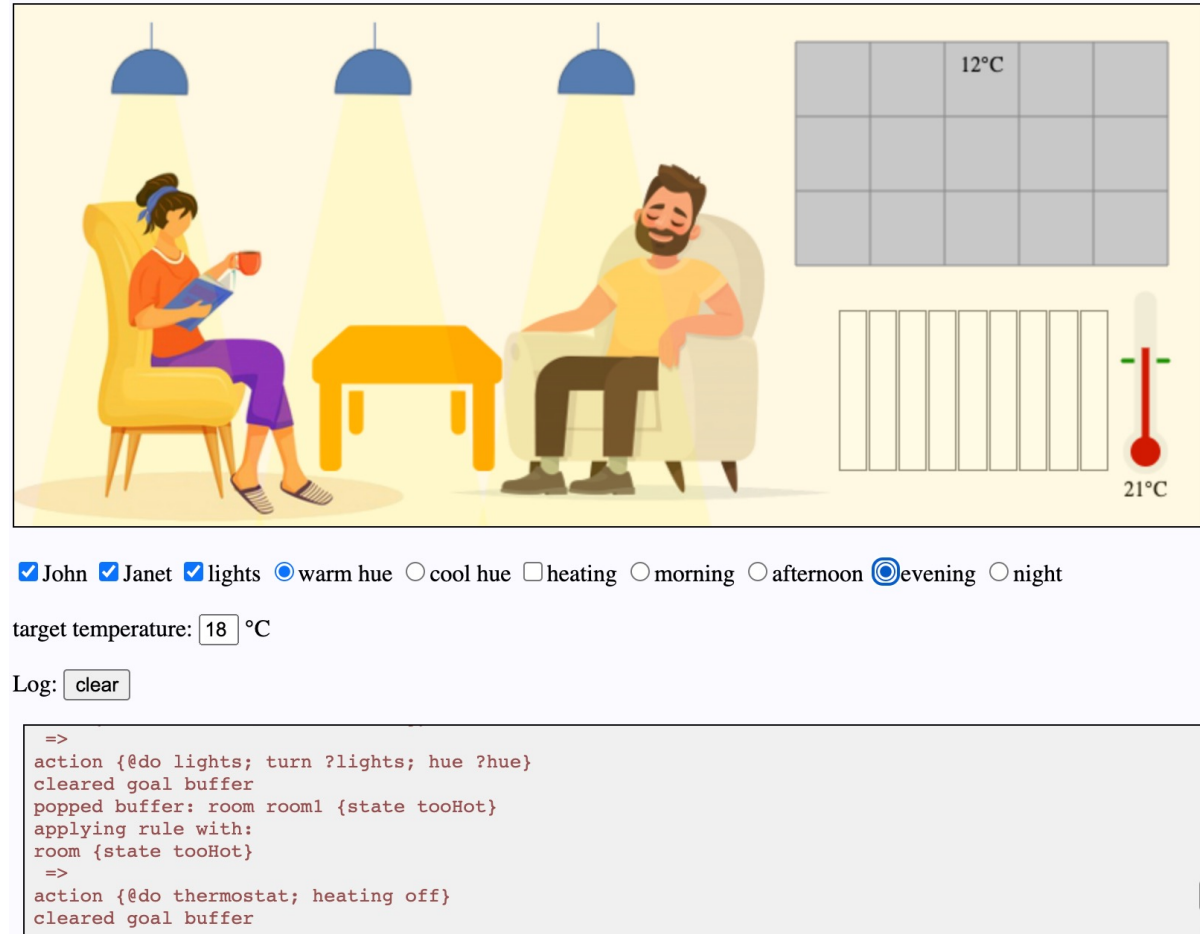
Log:

```
executed rule _:_:19 stop
set goal to: after _:_:54 {step 1}
executed rule _:_:27 move
set goal to: after _:_:55 {step 2}
executed rule _:_:30 grasp
set goal to: after _:_:56 {step 3}
starting belt1
wait on filled
executed rule _:_:34 start
```


Default Reasoning – Smart Home

<https://www.w3.org/Data/demos/chunks/home/>

- Dynamic simulation of smart home
 - Live thermal model of heat flows between home and outside world
 - Control of lighting and heating
 - Manually
 - Automatically
 - Forms-based control of who is in the room, and the time of day
- Mix of declarative and procedural knowledge
 - Personal preferences and priorities
 - Example of default reasoning
- Web page with JavaScript library for Cognitive AI



The interface shows a living room scene with a woman reading and a man sitting. Three blue pendant lights are on, casting yellow light. A yellow coffee table is in the center. On the right, there is a grey grid representing a window with '12°C' in the top-right cell. Below the window is a white radiator and a red thermometer showing '21°C'. Below the scene is a control bar with checkboxes for 'John', 'Janet', and 'lights' (all checked), radio buttons for 'warm hue' (selected), 'cool hue', 'heating', 'morning', 'afternoon', 'evening' (selected), and 'night'. Below the control bar is a 'target temperature: 18 °C' field. Below that is a 'Log: clear' button. At the bottom is a code editor showing the following JavaScript code:

```
=>
action {@do lights; turn ?lights; hue ?hue}
cleared goal buffer
popped buffer: room room1 {state tooHot}
applying rule with:
room {state tooHot}
=>
action {@do thermostat; heating off}
cleared goal buffer
```

Warehouse – Swarm Simulation

- Multi-agent system
 - Autonomous mobile & static agents that communicate by exchanging messages
- Swarms can be smarter than their components
 - Emergent behaviour
 - Ant colony metaphor
- Smart agents using cognitive control
 - Flexible low-code framework using cognitive rules that operate on knowledge graphs
 - Local or cloud based
- Other agents using hard-coded behaviour
- User interface for high-level monitoring and control

See: <https://www.w3.org/Data/demos/chunks/warehouse/>



Pan: shift + left button + mouse move (or with a single pressed finger). Zoom: shift + mouse wheel (or with two fingers).

SimSwarm: multi-agent swarm simulator

This is a web-based simulation of autonomous static and mobile agents that form a coordinated swarm, using chunks and rules for cognitive control. The floor tiles are 10 metres apart. For background information, see [simswarm talk](#).

Dave Raggett <dsr@w3.org>

Log:

```
loaded 0 facts
loaded 0 rules
```

► Knowledge Graph:

```
# some declarative facts - currently only for explanatory purposes
```



Swarm Management



❑ Centralised Control

- ❑ Swarm coordinator allocates tasks, tracks where everything is, and computes routes for mobile agents
 - Centralised mapping service
- ❑ Messages exchanged between swarm participants and the coordinator (one to one)
- ❑ Single point of failure
- ❑ But simpler in respect to monitoring and control for business objectives

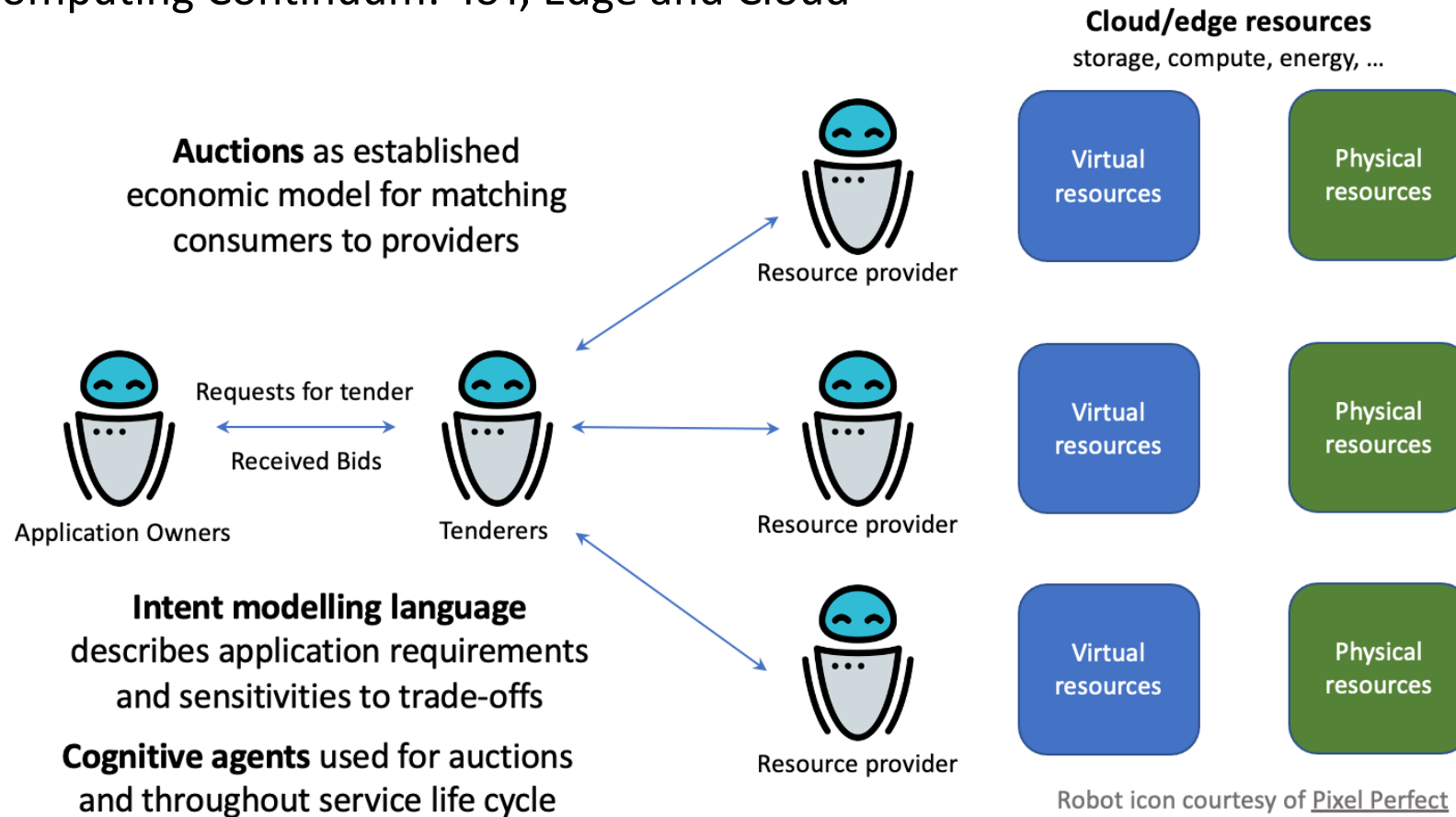
❑ Decentralised Control

- ❑ Tasks allocated in a distributed way:
 - 1) request for service, 2) offers of service, 3) requester notifies chosen task provider
 - Offers can specify a future time, e.g. when a forklift expects to finish its current job
- ❑ Topic based message distribution (one to many) – service providers listen on topics relevant to what services they provide
 - Can be limited to nearby participants
- ❑ Geospatial streams together with agents that offer routing services

Applying Cognitive Agents to Efficient Resource Management & Orchestration

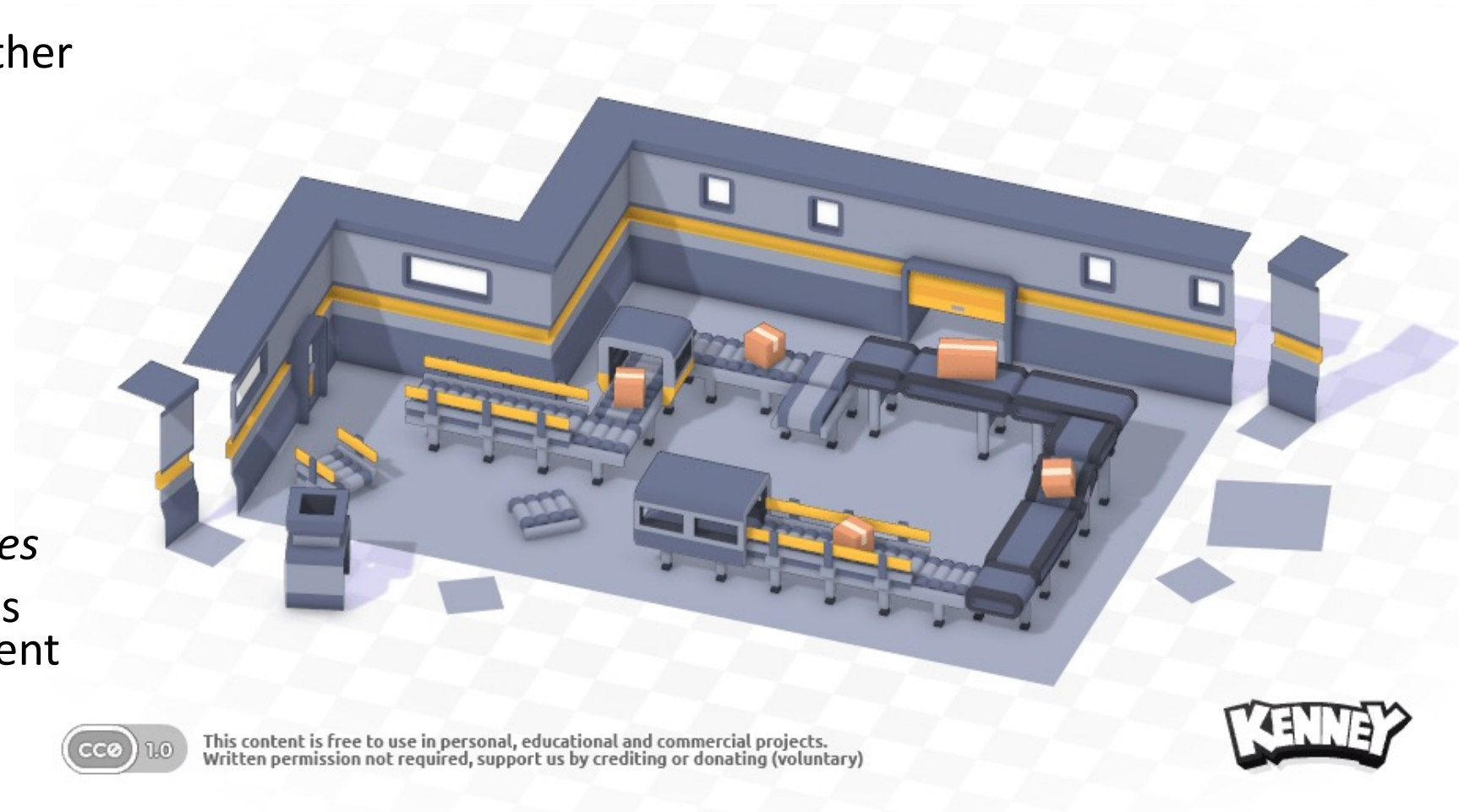
Across the Computing Continuum: IoT, Edge and Cloud

Cognitive agents
working on behalf
of their owners



Smart Factories

- Smart factories are another potential application
- Robot arms
- Milling machines
- Painting
- Conveyor belts
- Movable racks
- AMRs
- *and many other machines*
- People – as real factories use people to complement automation



This content is free to use in personal, educational and commercial projects.
Written permission not required, support us by crediting or donating (voluntary)

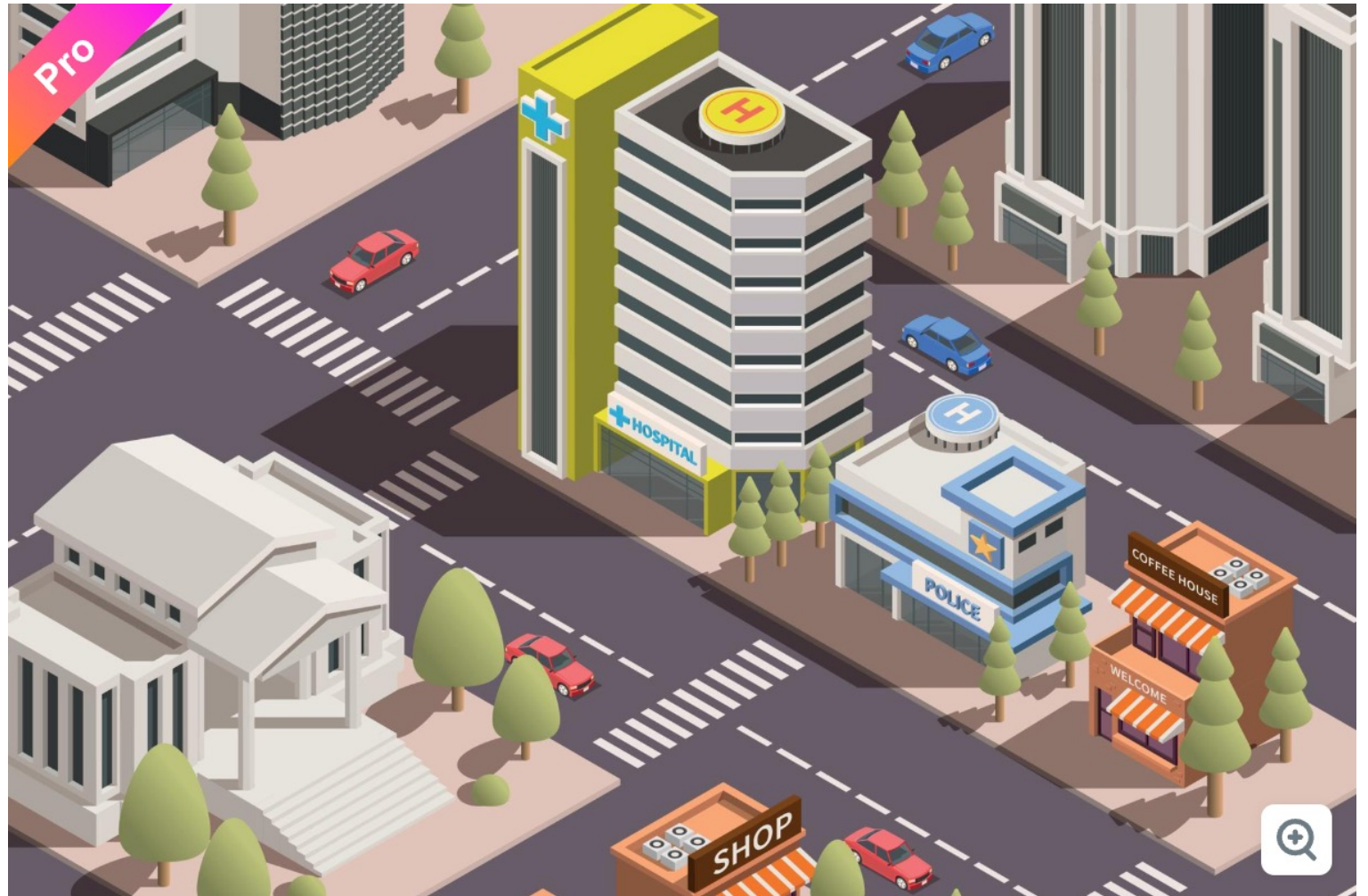


See: <https://kenney.nl/assets/conveyor-kit>

Urban Traffic

- Close up of just one type of road junction
- Challenges for large scale simulations*
- Could include cars, trucks, cyclists, and pedestrians, e.g. mothers with prams, old people with walking sticks, and young children

* Lower fidelity simulation outside field of view



Logistics

- International trade depends on efficient handing of containers
 - Transported on ships, trucks and trains
 - Minimising delays
 - Maximising throughput
- Warehouses for redistribution when containers carry loads for multiple customers
- Requirements for customs inspections and tariffs
- So called “Freeports”



Some Quotes on AI

- Marvin Minsky, 1967 (MIT media lab)

“AI will be substantially solved within a generation!”

- Microsoft Researchers on GPT-4, April 2023

“Some of the limitations of the next-word prediction paradigm, which manifest as the *model’s lack of planning, working memory, ability to backtrack, and reasoning abilities.*

The model relies on a local and greedy process of generating the next word, *without any global or deep understanding* of the task or the output.

Thus, the model is good at producing fluent and coherent texts, but has *limitations with regards to solving complex or creative problems* which cannot be approached in a sequential manner.”

- Yann Le Cun, May 2023, (Chief AI scientist for Meta)

“Machine learning sucks at least compared to humans and animals, which learn really quickly, using reasoning and planning.”

He says that autoregressive LLMs are doomed as they cannot be made factual, non-toxic, etc. and are not controllable. Errors in prediction accumulate so that the generated text irreversibly and exponentially diverges from the correct answers. He thinks that in five years time no one will use LLMs as these problems are recognised more widely.

- Andrew Ng , May 2023 (Stanford University, and a cofounder of Google Brain)

“I’m struggling to see how AI could pose any meaningful risks for our extinction. No doubt, AI has many risks like bias, unfairness, inaccurate outputs, job displacement, concentration of power. But I see AI’s impact as massively contributing to society, and I don’t see how it can lead to human extinction.”

- Yaser Abu-Mostafa, May 2023 (CalTech)

“AI now promises to alleviate routine mental labour and will take perhaps 20 years for widespread adoption. This is a very rapid timescale compared to the industrial revolution. We need a safety net to minimise the pain of the transition.”

He dismisses the risk of rogue AI systems.

“We attribute human intelligence to AI systems, and likewise our fears about bad people. It would be impractical for an AI system to develop a very much better system due to the challenges around building new chips, and assembling these into large scale systems. Studies show that general intelligence has no correlation with the will to dominate, hurt and control.”

He is concerned about social impact and increasing isolation. He says humans are very dependent on human contact.

World Economic Forum 2020: “new jobs will emerge and others will be displaced by a shift in the division of labour between humans and machines, which on the balance will boost the number of jobs available. This merits support for re-training programmes to help people benefit from the changes.”

Risks and Liabilities

- ❑ The European Parliament is reviewing draft regulations on handling AI Risks (EU AI Act)

- Designed to complement
 - Digital Services Act (DSA)
 - Digital Markets Act (DMA)
- Defines levels of risk for AI systems
 - unacceptable, high, limited and low

- ❑ Further legislation expected on Liability Rules for AI

- ❑ Unacceptable risk

All AI systems deemed a clear threat to the safety, livelihoods and rights of people will be banned

- ❑ High risk

Subject to extensive obligations throughout their lifecycle, e.g. for risk assessment and mitigation

- ❑ Limited risk

Users must be made aware that they are interacting with an automated system

- ❑ Low risk

No restrictions

Closing Thoughts

- ❑ AI is rapidly improving but we are still a long way from AGI*
- ❑ Neural networks automate knowledge engineering
 - Especially for unstructured data, e.g. text, images, sound and video
- ❑ But symbolic AI remains crucial for semantic interoperability
 - Emergence of cognitive databases
- ❑ Large language models can play a collaborative role for developing and maintaining ontologies
- ❑ AI has a huge potential to increase prosperity* despite the naysayers
 - boosting employment and wages
- ❑ Open source AI is a great way to speed research and exploitation
- ❑ AI should be applied to helping address societal challenges, e.g.
 - misinformation, circular economy, malnutrition, disease and climate
- ❑ Competing visions
 - AI for open liberal societies vs a tool for authoritarian population control

* Artificial General Intelligence, i.e. sentient minds

* [Why AI will save the world](#), Marc Andreessen

Questions and Comments?



This work has been supported through funding for the TERMINET project from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957406.