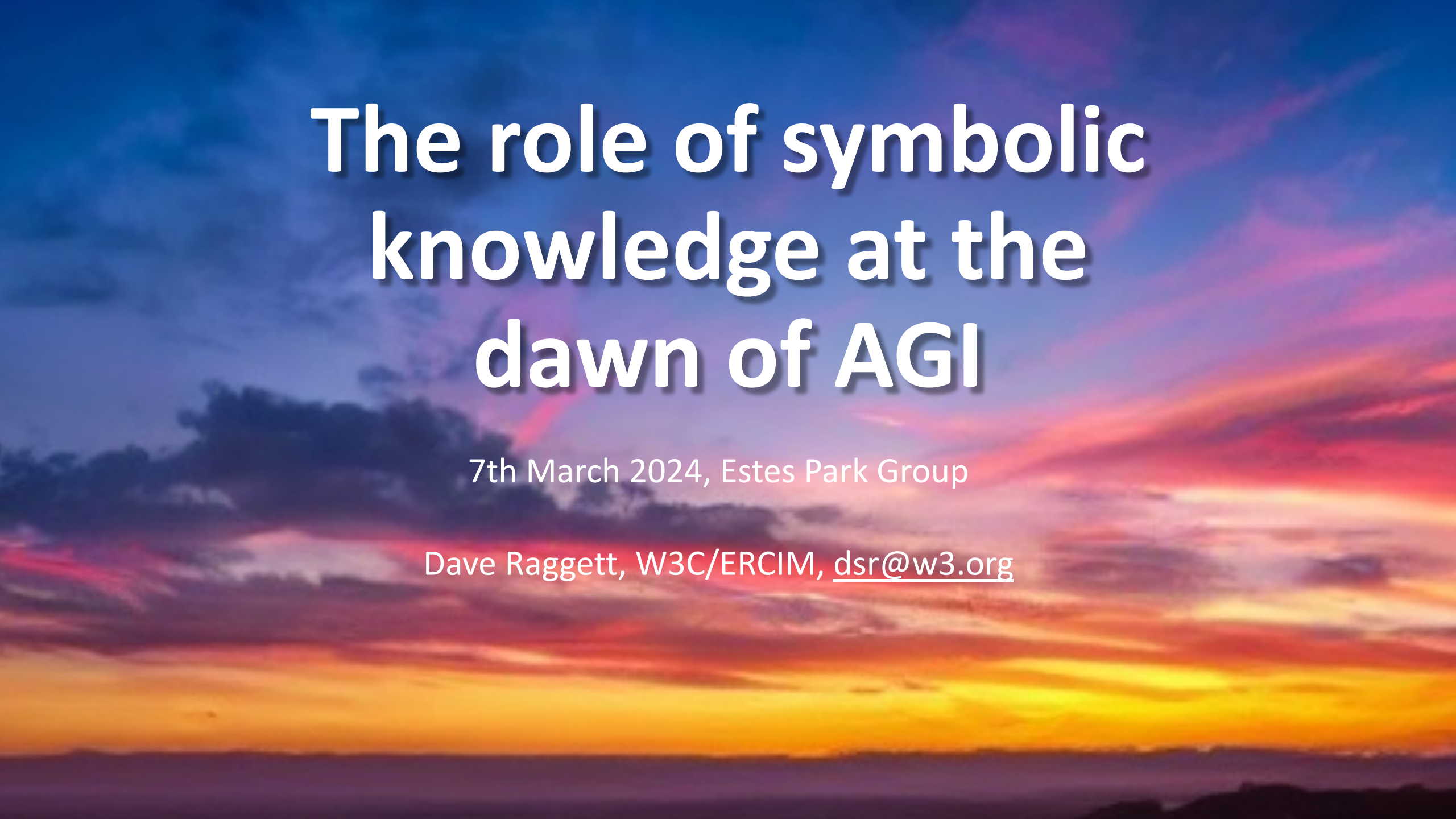


The role of symbolic knowledge at the dawn of AGI

7th March 2024, Estes Park Group

Dave Raggett, W3C/ERCIM, dsr@w3.org

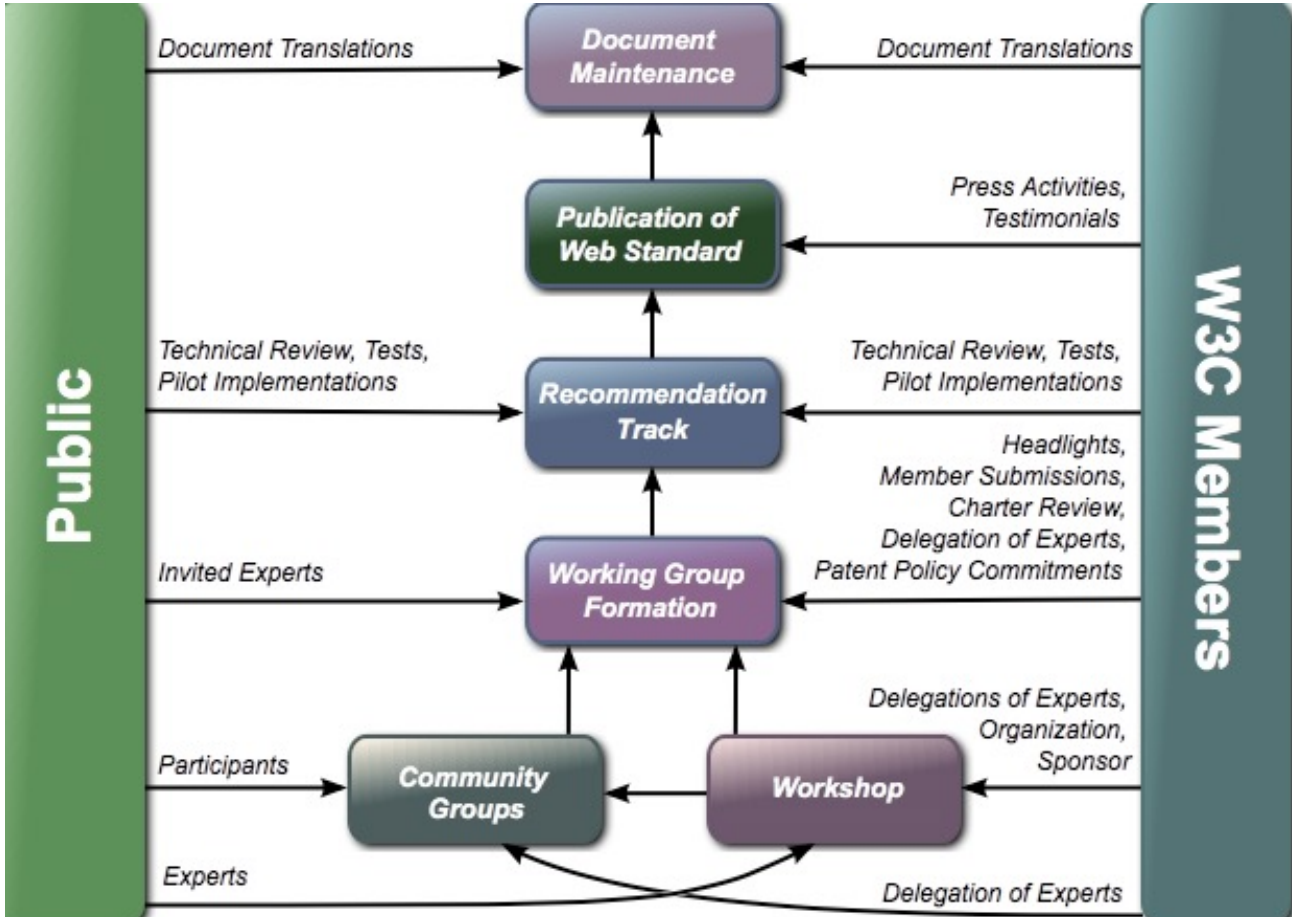
Table of Contents

Generated with DALL-E 3

World Wide Web Consortium*

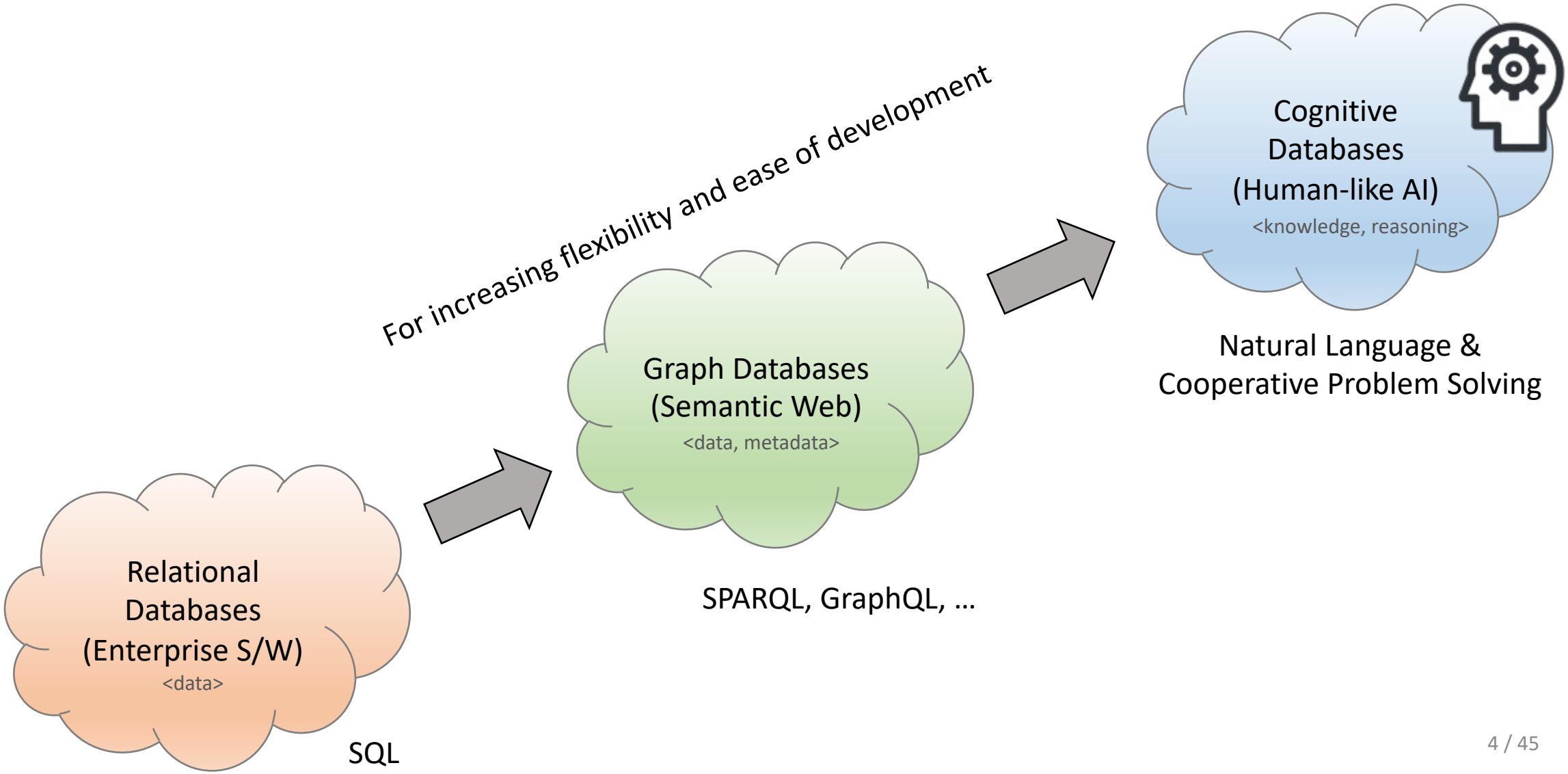
www.w3.org

- International member-funded community working on open standards for the Web since 1994
- Focus on interoperability for web browsers and websites, including linked data, the semantic web and the web of things
- 7,500 specifications including 440 W3C Recommendations
- Enabling people with disabilities to access the Web
- Built-in support for many of the world's languages
- A ground-breaking royalty-free Patent Policy



* I work for ERCIM, the European partner for W3C.org. ERCIM is the European Research Consortium for Informatics and Mathematics. We aim to foster collaborative work within the European research community and to increase co-operation with European industry.

Evolution in ICT Systems



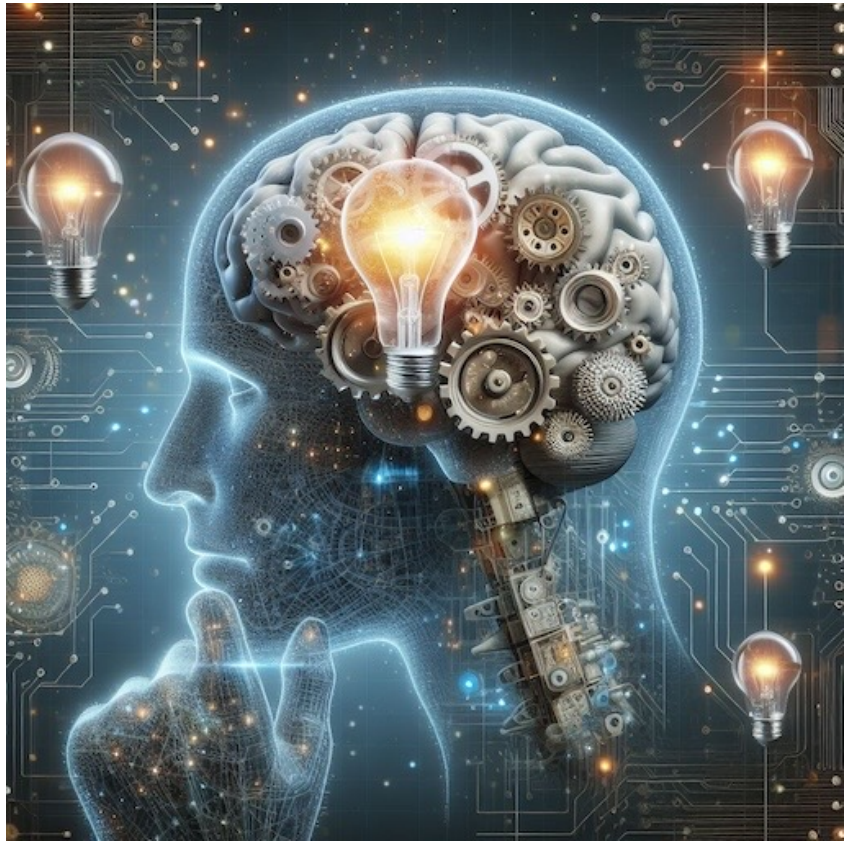
Limitations of Logic and Deductive Proof

- **Logic** deals with mathematical entailments of what is held to be true
- This assumes perfect unchanging knowledge
- Logic isn't applicable for knowledge that is uncertain, context sensitive, imprecise, incomplete, inconsistent and changing, i.e. imperfect knowledge
- That is however typically the case for everyday knowledge
- **Defeasible Reasoning** is much broader than logic and forms the basis for legal arguments, ethics, political arguments and everyday discussions
- We should embrace the challenge!

Reasoning has been studied since the days of Ancient Greece

Defeasible Reasoning and Argumentation

- **Deductive proof** is replaced by **defeasible reasoning** with arguments for and against suppositions
- Strict rules logically entail their conclusions, whilst defeasible rules create a presumption in favour of their conclusions, which may need to be withdrawn in the light of new information
- Arguments in support of, or counter to, some supposition, build upon the facts in the knowledge graph or the conclusions of previous arguments
- Preferences between arguments are derived from preferences between rules with additional considerations in respect to consistency
- Counter arguments can be classified into three groups. An argument can:
  - **undermine** another argument when the conclusions of the former contradict premises of the latter
  - **undercut** another argument by casting doubt on the link between the premises and conclusions of the latter argument
  - **rebut** another argument when their respective conclusions can be shown to be contradictory

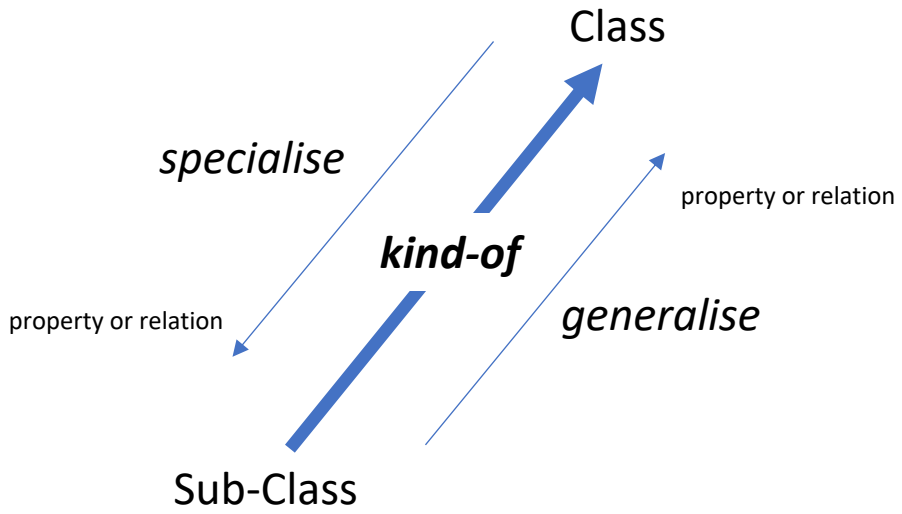


Argumentation Theory

- The Stanford Encyclopaedia of Philosophy lists five types of arguments: *deduction, induction, abduction, analogy* and *fallacies*
- Studies of argumentation have been made by a long line of philosophers dating back to Ancient Greece, e.g., Carneades and Aristotle
- More recently, logicians such as Frege, Hilbert and Russell were primarily interested in mathematical reasoning and argumentation
- Stephen Toulmin subsequently criticized the presumption that arguments should be formulated in purely formal deductive terms
- Douglas Walton extended tools from formal logic to cover a wider range of arguments – a set of argument schemes
- Ulrike Hahn, Mike Oaksford and others applied Bayesian techniques to reasoning and argumentation
- AIF is an ontology intended to serve as the basis for an interlingua between different argumentation formats
- Alan Collins applied a more intuitive approach to plausible reasoning that takes sub-symbolic knowledge into account to model rough notions of metadata in lieu of statistics
- Collins inspired my work on the **Plausible Knowledge Notation (PKN**) in the W3C Cognitive AI Community Group
- Arguments in support of, or counter to, some supposition, build upon the facts in the knowledge graph or the conclusions of previous arguments
- Preferences between arguments are derived from preferences between rules with additional considerations in respect to consistency

Plausible Inferences using Prior Knowledge

- Inferring likely properties and relations across other relations



- Expected certainty influenced by qualitative metadata
  - e.g. typicality, similarity, strength, dominance, multiplicity, scope, …
- Forward and backward inferences using implications
  - Given: *If it is raining then it is cloudy*
  - Infer: *If it is cloudy it may be rainy*
- Inferences based upon analogies
  - matching structural relationships
- Scalar ranges (fuzzy logic)
  - fuzzy terms, e.g. *cold, warm* and *hot*
  - fuzzy modifiers, e.g. *very* old
  - fuzzy quantifiers, e.g. *few* people …
- Multiple lines of argument for and against the premise in question

PKN Demonstrator

- Proof of concept implementation in JavaScript as a web page
- Large collection of examples
- Works back from the supposition towards the supporting facts
- Avoids circular arguments
- Explanation generated in forward pass through trace of execution

https://www.w3.org/Data/demos/chunks/reasoning/

www.w3.org/Data/demos/chunks/reasoning/

Search Bing

Use the drop-down menu below to select which query to reason about. Use the *effort* checkbox to seek indirect evidence even when direct evidence is found, and the *trace* checkbox to see reasoning in action in addition to the explanation generated from it.

Whether daffodils are grown in England? Next Previous

Effort: seek additional evidence. Trace: more details.

```
Premise: flowers of England includes daffodils
Evidence supporting the premise:

  flowers of England includes temperate-flowers (certainty high)
  and daffodils kind-of temperate-flowers
  therefore flowers of England includes daffodils (certainty high)

  flowers of Netherlands includes daffodils,tulips (certainty high)
  and Netherlands similar-to England for flowers
  therefore flowers of England includes daffodils (certainty high)

Suggesting it is likely that flowers of England includes daffodils (certainty high)

No evidence found that flowers of England excludes daffodils (certainty high)
```

▼ Plausible Knowledge graph:

```
# Example Plausible Knowledge Graph

# a simple taxonomy
daffodils kind-of temperate-flowers
tulips kind-of temperate-flowers
roses kind-of temperate-flowers
temperate-flowers kind-of flowers
flowers kind-of plants

# used to infer that daffodils grow in England
flowers of England includes temperate-flowers
flowers of Netherlands includes daffodils, tulips
flowers of Netherlands includes roses
Netherlands similar-to England for flowers

# used to infer climate of England
Netherlands similar-to England for climate
climate of Netherlands includes temperate

# used to infer climate of Belgium
Belgium similar-to Netherlands for latitude
climate depends-on latitude

# example of conflicting knowledge
```

higher level than RDF

PKN Examples

*The Plausible Knowledge Notation (**PKN**) includes enriched semantics and an easier to use notation relative to RDF/turtle*

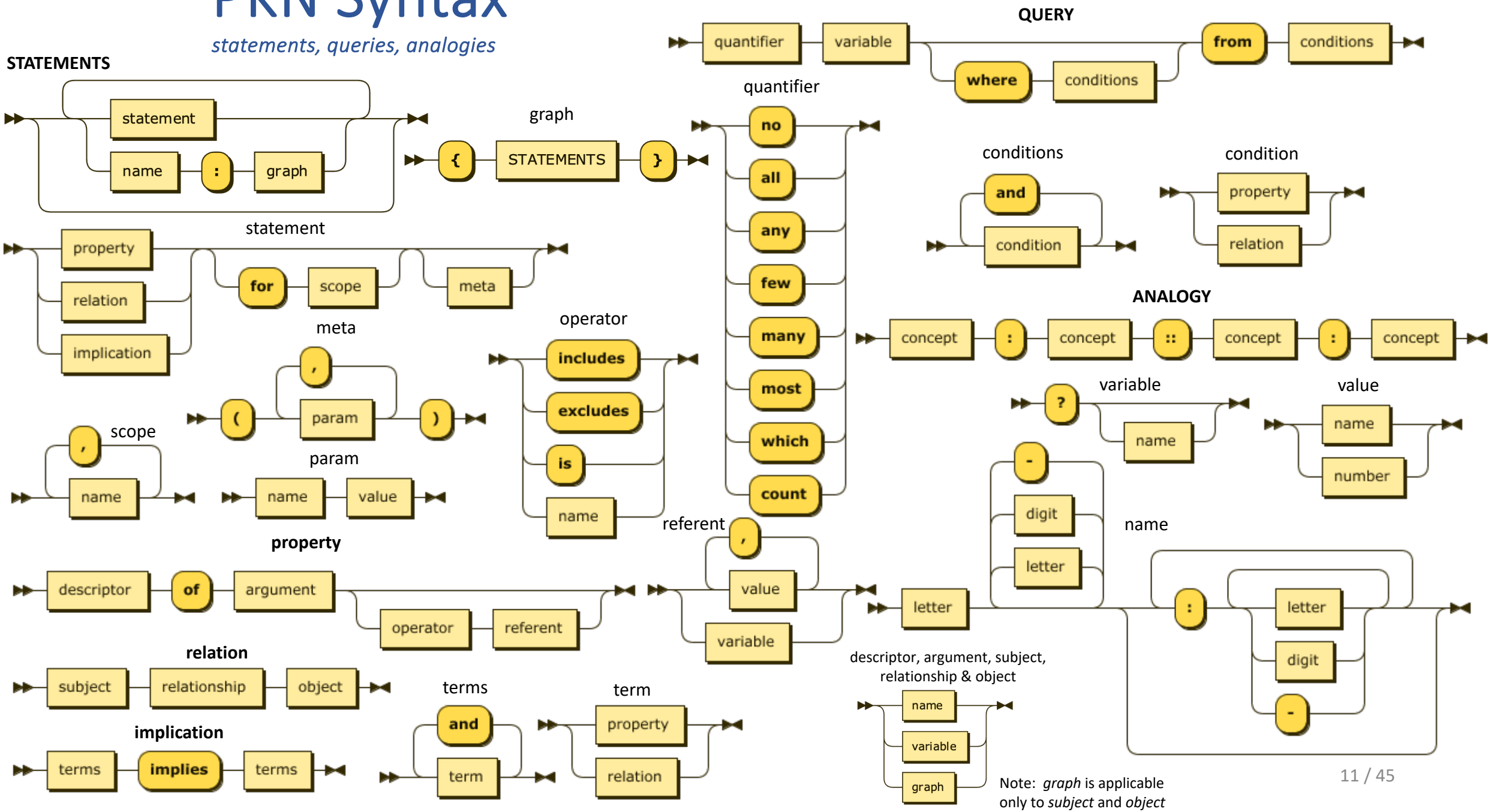
properties, relationships, contextual scope, implication rules, fuzzy ranges, fuzzy modifiers, fuzzy quantifiers, analogies, parameters denoting gut feelings, statements about statements

See: PKN specification and Web based demo

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
flow:current::pressure:voltage
dog:puppy::cat:?
weather of ?place includes rainy
    implies weather of ?place includes cloudy (strength high, inverse low)
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)
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
Mary believes {{John says {John loves Joan}} is-a lie}

PKN Syntax

statements, queries, analogies



Further Details

- Context dependent relations
  - Belgium similar-to Netherlands for latitude
- Fuzzy ranges and context sensitivity
  - Defining *age* with a set of fuzzy terms that depend on whether you are a child or an adult
- Fuzzy modifiers
  - Paul close:friend-of John
- Fuzzy quantifiers using set comparisons
  - few ?x where color of ?x includes yellow from ?x kind-of rose
- Imagination – planning , what-if analysis, understanding intent, modelling stories, reported speech
  - Named and unnamed collections of statements
  - which ?x where Joan said {?x likes tea}

- Reasoning over related ontologies
  - Different terminologies for climate
    - dry, temperate, continental and polar
    - polar, temperate, arid, tropical, mediterranean, mountain
    - Hot: equatorial, tropical, subtropical; Temperate: Mediterranean, chinese, oceanic, continental; Cold: polar, highland
  - Each of these terms is associated with typical weather patterns for that climate, e.g. in cities like Shanghai, Buenos Aires, Sydney, and Hong Kong which have a so called Chinese climate with mild winters and humid summers with tropical rain
  - This points to the potential for using defeasible reasoning as there is no one way to relate the terms, and it is more about searching for *support-for* or *counter-to* the supposition in question with varying degrees of certainty

For more details, see: PKN specification

Strategies and Tactics for Argumentation

- Further work is now needed on an intuitive syntax for *reasoning strategies and tactics*, as well as ways to model *pathos* – i.e. the role of feelings and emotions as part of compelling arguments
- Building upon well established **principles** for **effective arguments**, e.g. classical rhetorical guidelines dating back to Aristotle
  - **Ethos**: establishing credibility to engender trust
  - **Pathos**: using emotion to stir people's feelings
  - **Logos**: using logic to emphasise rational support
  - **Kairos**: opportune, i.e. timely and topical in nature
  - Use of rhetorical questions to strengthen support
- Need to gather use cases and suite of examples

Cognitive AI

- Human intelligence
- Thinking & Problem Solving
- Memory
- Learning
- Language
- Perception
- Actuation
- Attention
- Feelings and emotions

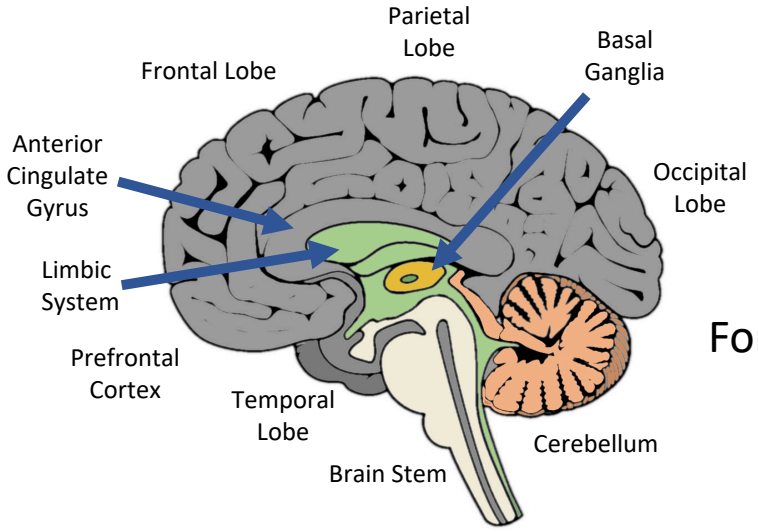
What the Cognitive Sciences can tell us

- The interdisciplinary study of the mind and its processes
  - linguistics, psychology, neuroscience, philosophy, AI and anthropology
- Decades of work in the cognitive sciences on understanding the mind, how we learn, the kinds of mistakes we make, …
- This can provide deep insights for working with neural networks
- Mixing symbolic and sub-symbolic models
  - John Anderson on ACT-R with chunks and rules
  - Alan Collins on plausible reasoning
  - Dedre Gentner on analogical reasoning
  - Lotfi Zadeh on fuzzy reasoning
  - George Lackoff & Mark Johnson on role of metaphors

Human Language Processing is Sequential, Hierarchical and Predictive

- Evidence from:
  - Eye saccades when reading text
  - Buffering limitations for phonological loop
    - Few words *not* thousands of words
  - Semantic priming effects
    - Word sense disambiguation based upon previous and following words
  - Brains scans for active areas
- Bottom-up processing for sounds, and syllables before words and sentences
  - Sequential with limited overlapped processing
- Top-down using the context and prior knowledge
- Processing is both hierarchical and predictive
  - Incremental learning of new vocabulary

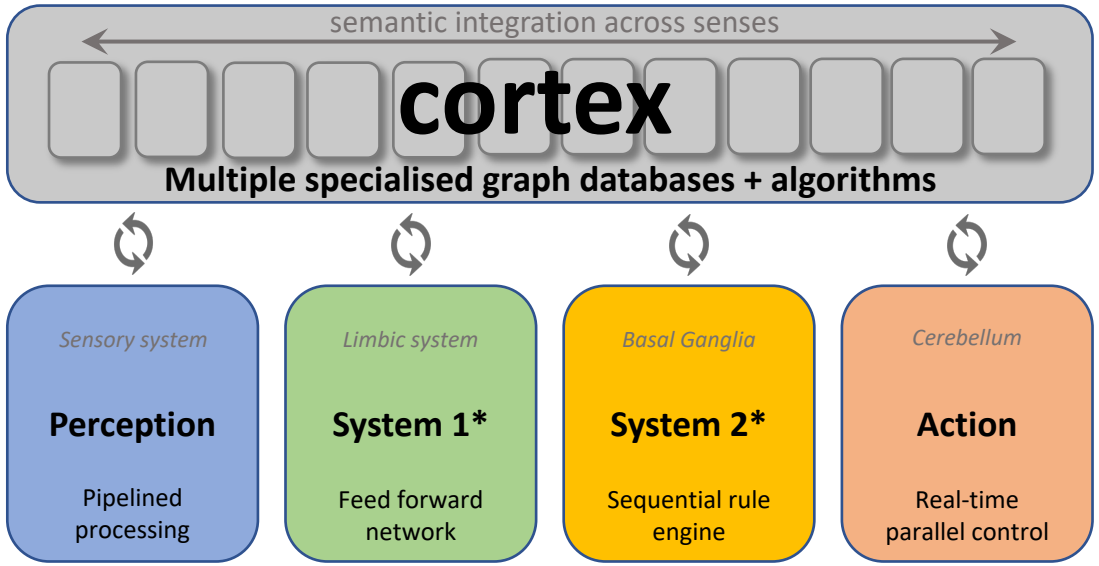
See also: Human-like systematic generalization through a meta-learning neural network (October 2023)



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 for semantic priming, the forgetting curve and spacing effect. Hub and spoke model is used 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, e.g. driving a car, 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.

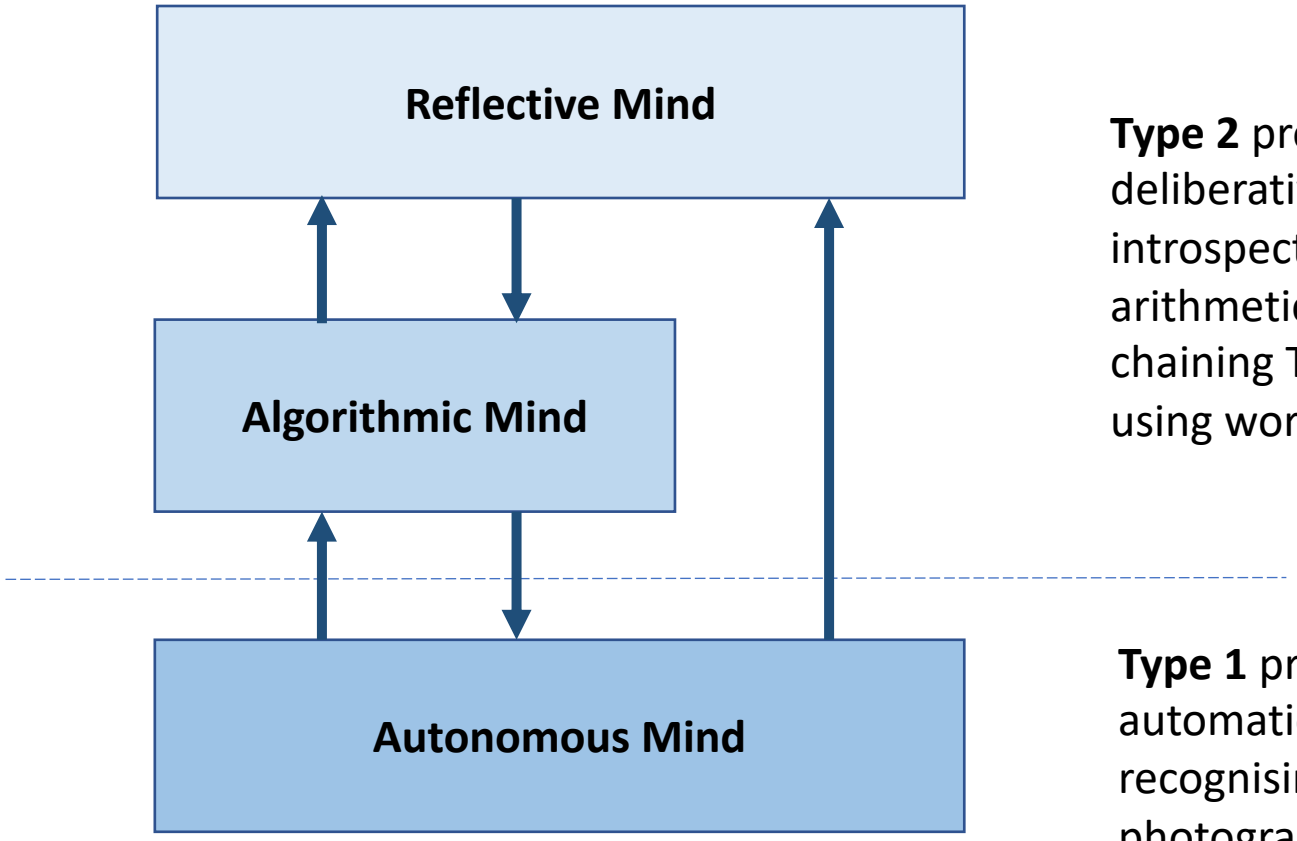
System 1 covers intuitive/emotional thought, cognitive control and prioritising what's important. The limbic system provides rapid assessment of past, present and imagined situations. Emotions are perceived as positive or negative, and associated with passive or active responses, involving actual or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

* You will also see the terms Type 1 and 2 processing

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.

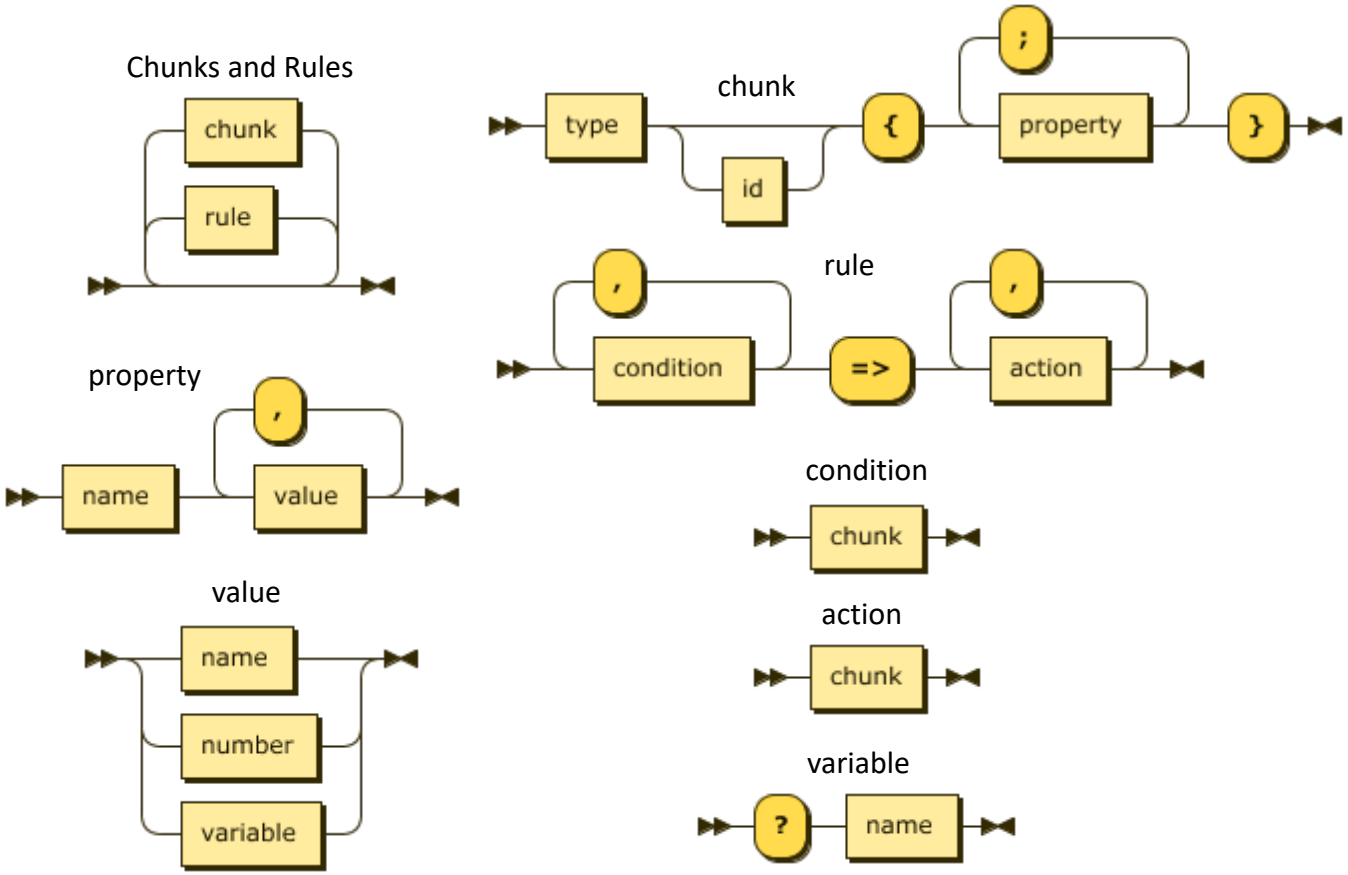
See, e.g. "Dual-Process Theories of Higher Cognition: Advancing the Debate", Evans and Stanovich (2013), along with "Thinking Fast and Slow", Daniel Kahneman (2011)

Chunks and Rules*

web-based demos for smart homes and factories

higher level than RDF

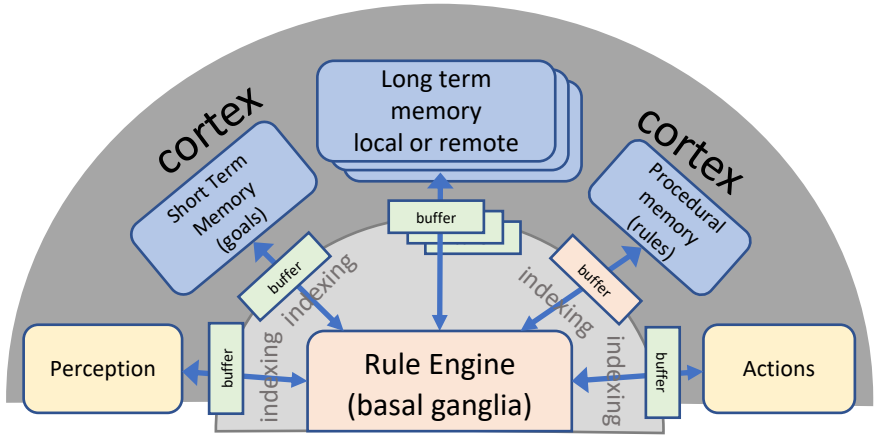
```
# move robot arm into position to grasp empty bottle
after {step 1} =>
    robot {@do move; x -170; y -75; angle -180; gap 30; step 2}
```



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 request-response model

- Inspired by John Anderson's ACT-R and decades of cognitive science research at CMU and elsewhere
- 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 built-in cortical operations
- Reasoning decoupled from real-time control over external actions, e.g. a robot arm

Limitations of Symbolic AI

- Symbolic AI is generally hand-crafted
- Impoverished representations compared to the subtle context sensitivity and imprecision common in the real-world
- This results in problems in practical use
- Expensive to develop and maintain - as a result **very hard to scale up**
- Recent successes for Generative AI show that computers can be very much better at knowledge engineering than we are
- **Are we wasting our time on symbolic AI?**

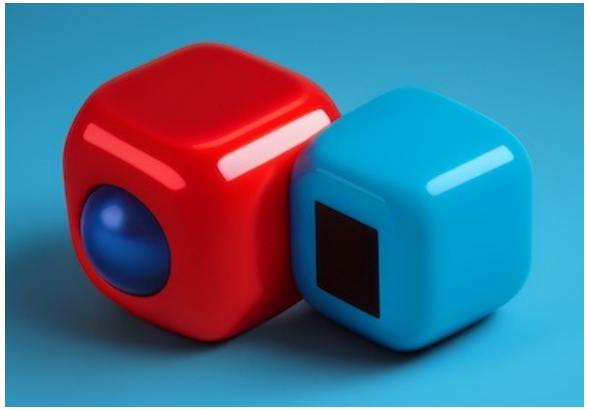
Limitations of today's Generative AI

Generated using DeciDiffusion

Generative AI

- Astonishing ability to learn billions of parameters in complex neural networks via back propagation
- Amazing capabilities in dealing with text and images, and now being extended to speech, music, video and 3D
  - *Many opportunities for multimodal applications**
- Chain of thought plus reinforcement learning with human feedback – success at passing our exams!
  - *Fine tuning and other techniques for ensuring safe responses, e.g. bootstrapping using self-critique from a set of principles*
- Prompt engineering as a valuable new skill!
  - *But LLMs will be able to craft good prompts for us*
- Prone to bias, distractions and hallucinations
- Weak on logical reasoning and semantic consistency
- Lack of continual learning and temporal memory
- Very expensive to train foundation models
- Very different from the human brain
- More like alchemy than science – *but early days yet!*

Hmm, how many fingers do humans have?



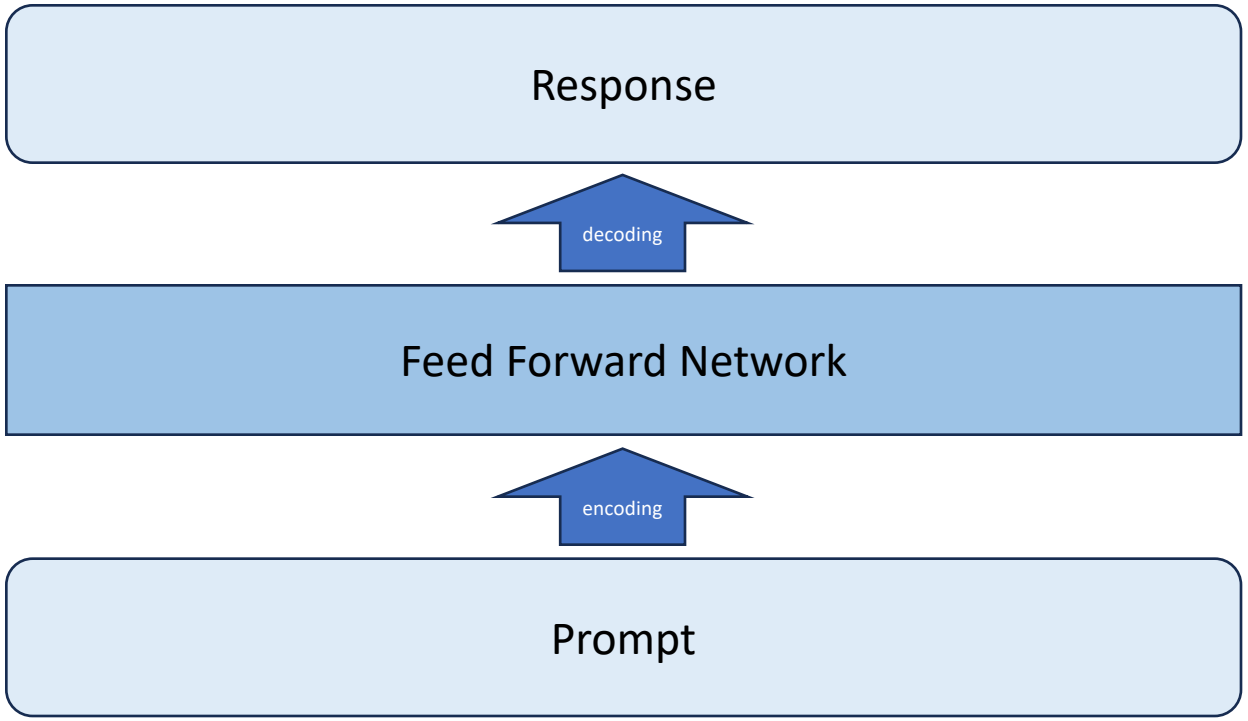
"3 red balls and 2 blue cubes on a wooden floor", **really???**

Is 1 kg heavier than 2 kg: **no** ✓

Is 1 kg of lead heavier than 2 kg of feathers: **yes** ✗

* Training on TV shows and Video will enable learning emotional models

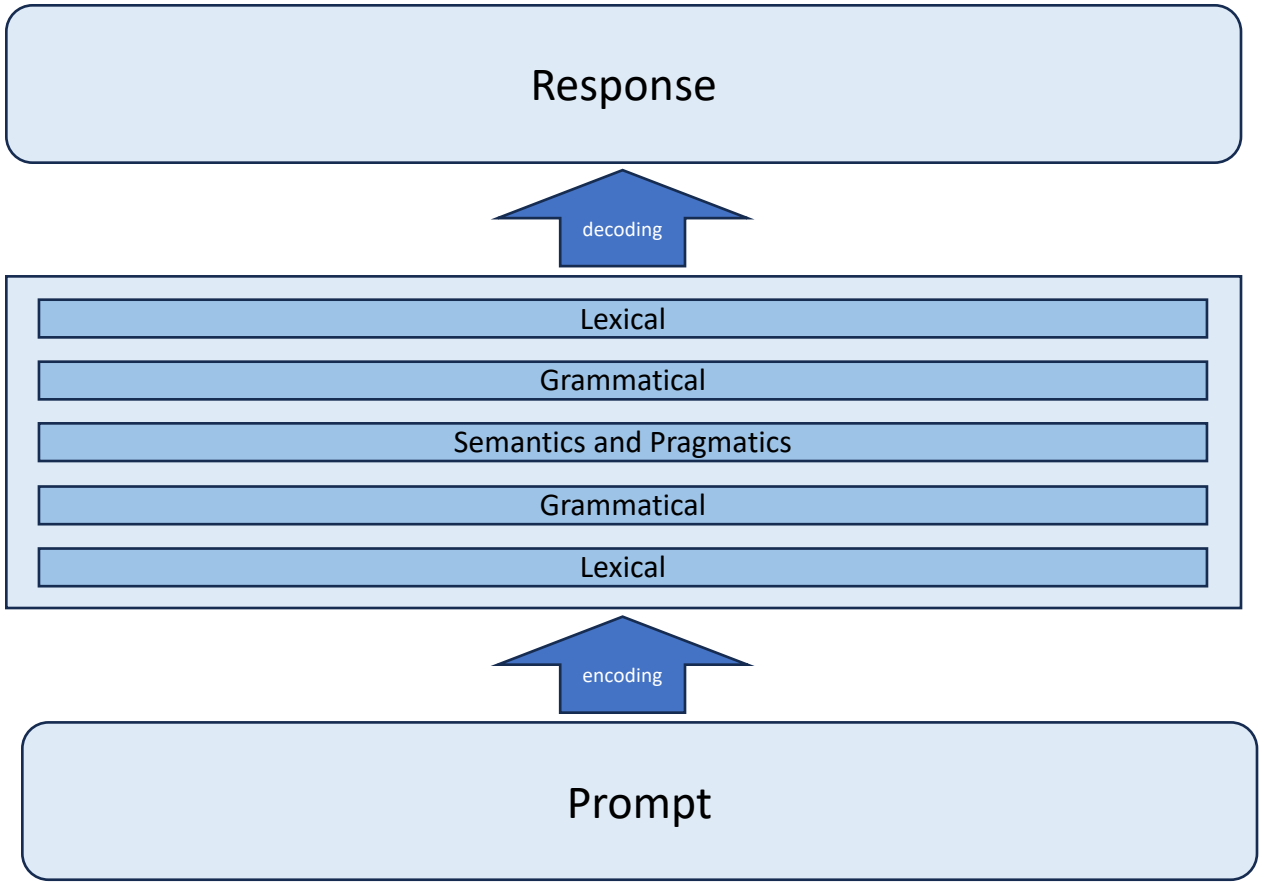
Large Language Models



- Neural network is used for **statistical prediction** of the response for the given prompt
- Text is encoded as sequence of tokens that are **vectors** in a text embedding space
- Feed forward network uses multiple layers of **Transformers** for long range attention and hierarchical dependencies
- Network params trained via **back propagation** on an loss function based upon text prediction of masked tokens
- **No short term memory**
- **No continual learning**

LLMs typically process thousands of text tokens in parallel which requires vast computing power

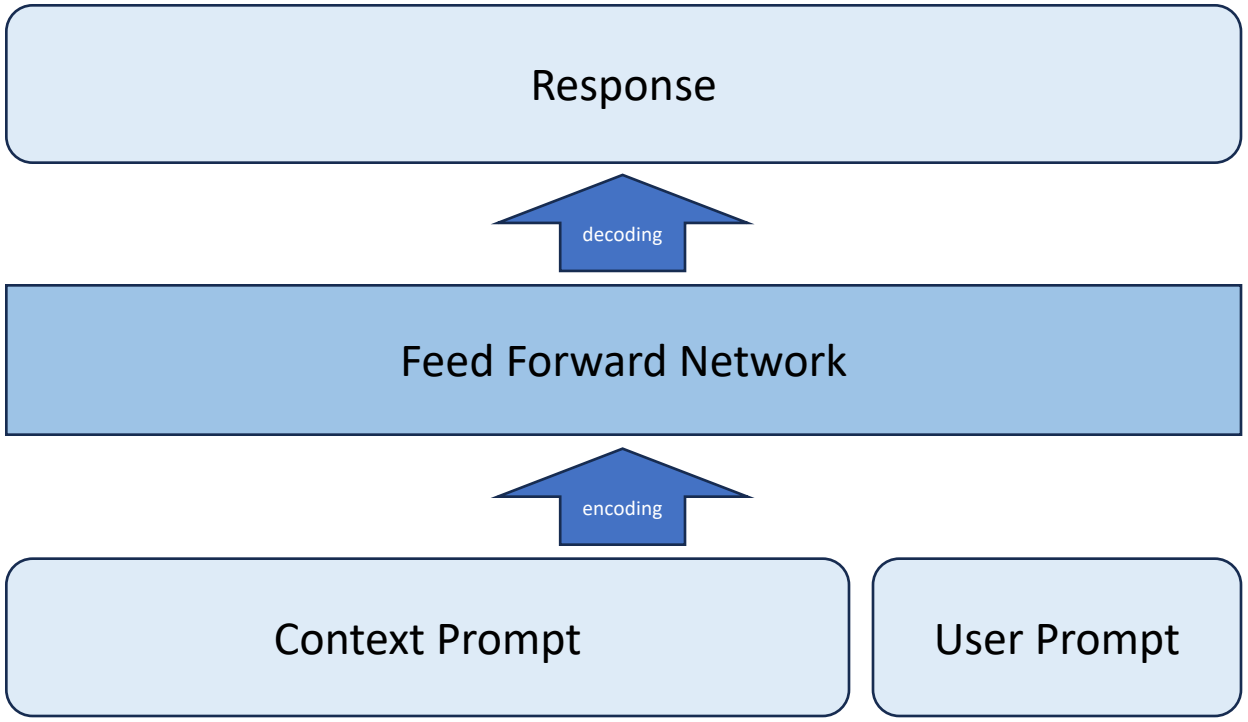
Latent Semantics Deep within the Network



- LLMs use neural networks with many richly connected feed-forward layers
- Network connections encode knowledge in a distributed fashion using vectors rather than symbols
  - parts of speech, word senses, grammatical structures, slot fillers, semantics and implicatures
  - opaque representations of knowledge
- Reliant on attention as a surrogate for reverse flow of information, e.g. from semantics to word senses
  - semantics implicit in nearby words and words that act as verb slot fillers, etc.
- Top and bottom layers closely related to word tokens
- Middle layers related to semantics and pragmatics

Pragmatics is the study of how context contributes to meaning including deixis, turn taking, text organisation, presupposition and implicature

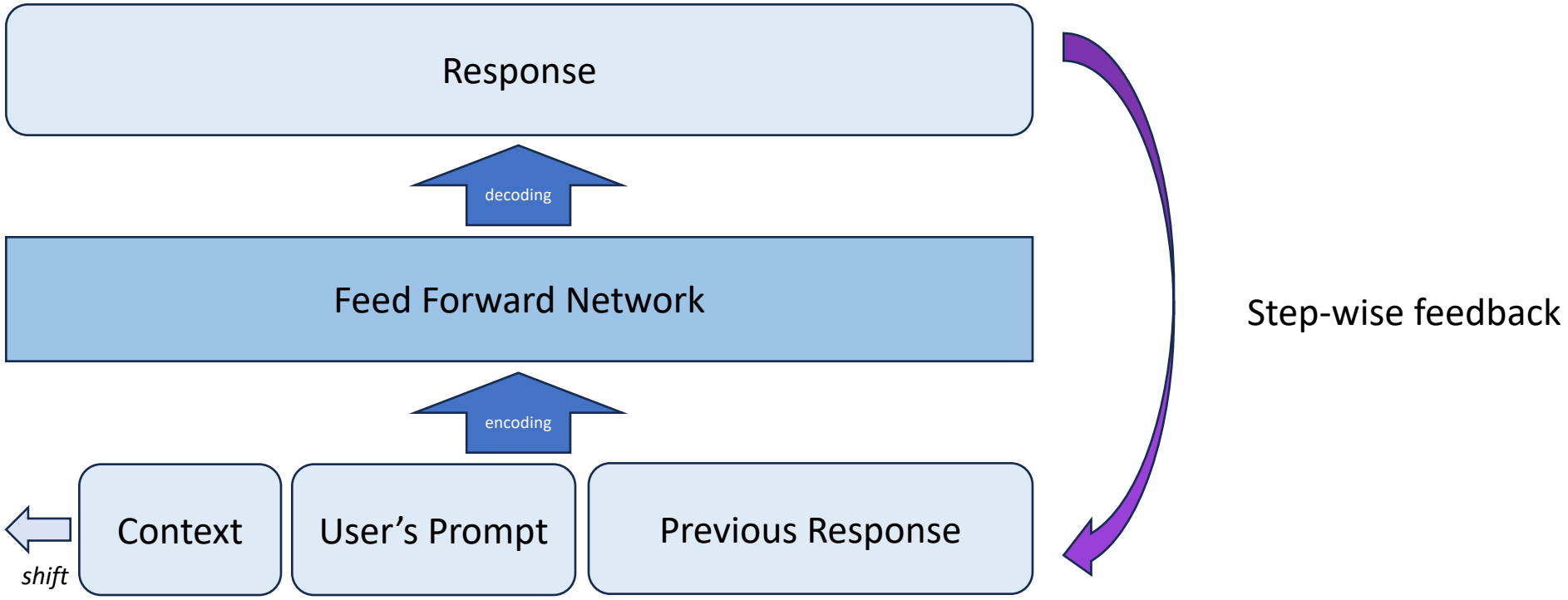
Tailoring the Context for User Prompts



Letting the LLM know what we're hoping for in the response

The context prompt is automatically injected to guide the kind of response to fulfil application requirements

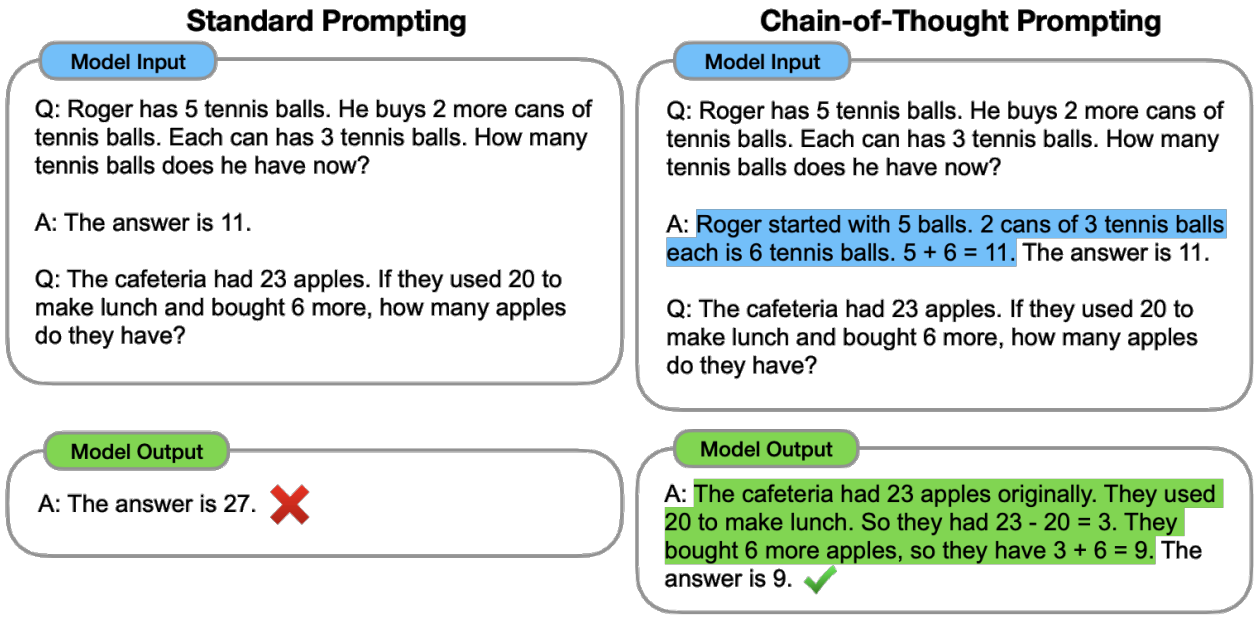
Feedback as a Surrogate for Short Term Dialogue Memory



The Response is appended to the Prompt as it is generated to provide a kind of short term memory, making it possible to generate lengthy responses

Prompt Engineering

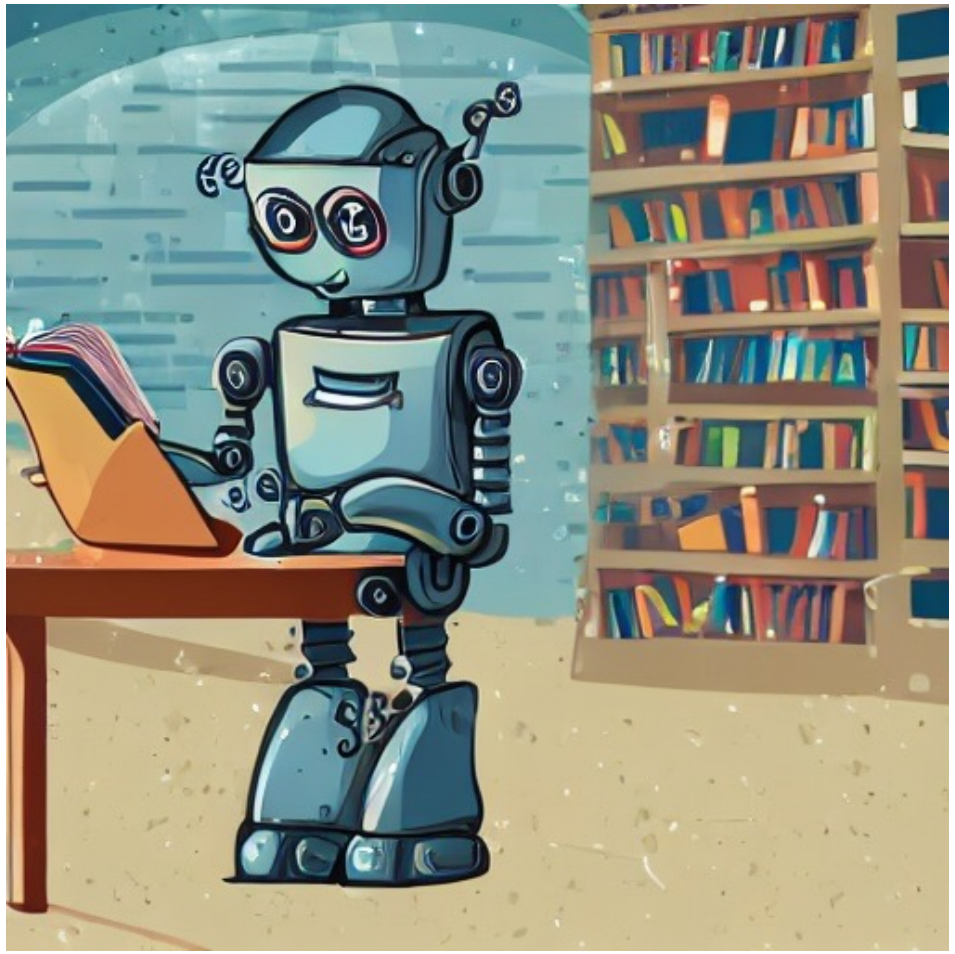
- Good prompts give good responses
- Different kinds of prompts, e.g.
  - Text completion, instruction-based, multiple-choice, least to most, search based, contextual, bias mitigation, chain of thought, tree of thought, …
- Generally speaking, specify what you want, e.g.
  - *Each title should be between two and five words long*
  - And provide a few examples as a guide
- Chain of thought prompting* to elicit sequential reasoning
  - Using worked examples
  - Improve results for specific domain via reinforcement learning with human feedback
- Adversarial attacks with crafted prompts
  - Bypassing LLM safety measures
- LLMs can be trained to craft expert prompts using our guidance to generate artwork or reports
  - Yang et al. Large language models as optimisers - OPRO (Sep' 2023), and try using ChatGPT via Bing search to generate prompts for DALL-E 3,



* Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, Jan' 2022

Retrieval Augmented Generation

- LLMs are trained once, and as such their knowledge is static
- Retraining LLMs is very expensive
- LLMs also have difficulties in generating citations for static knowledge embedded in their network parameters
- A work around is to query a knowledge graph to obtain a list of relevant sources and citations
- Then inject this as part of the context for the prompt and instruct LLM to generate links
  - Allows for up-to-date information and avoids need for LLM to include sensitive data
- Vector databases including text, images, …
  - dense vector index for external data acting on user's query to fetch most relevant citations



What are we looking for in AGI?

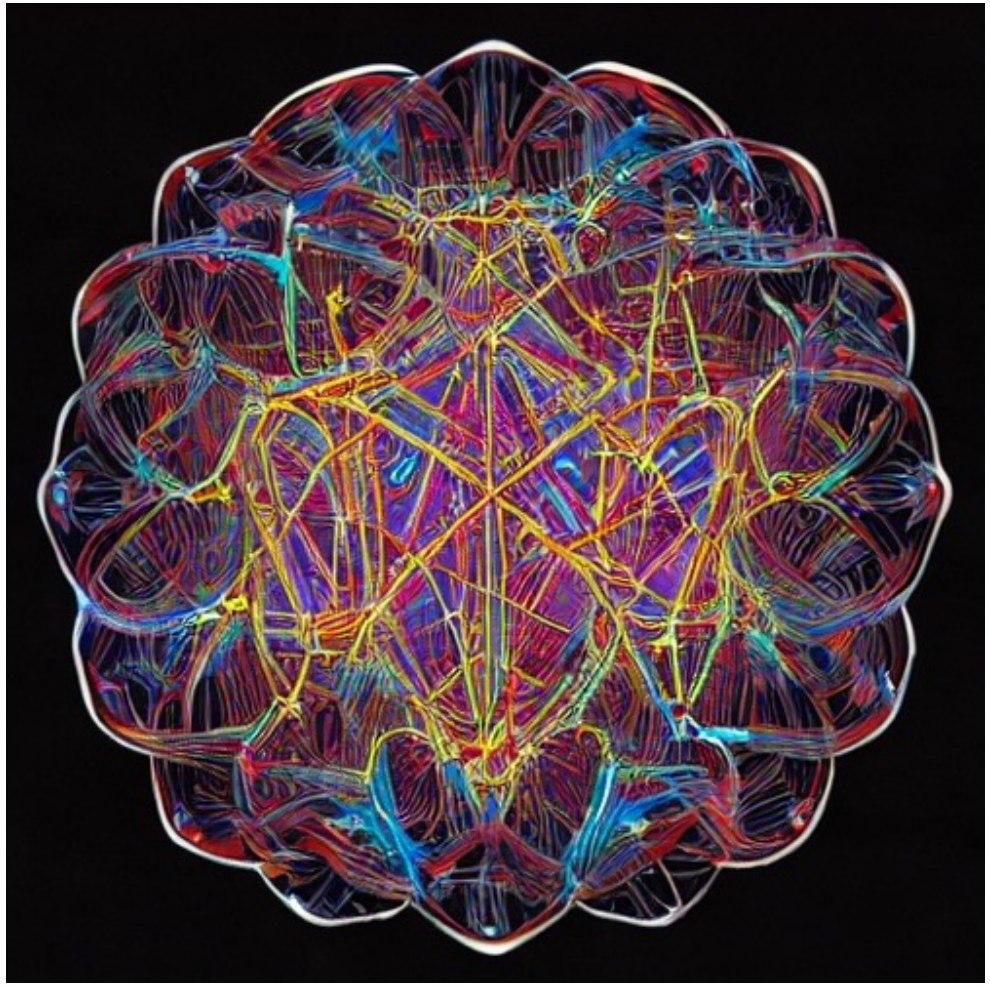
Artificial General Intelligence

- Creativity in problem solving, with the ability to generate and adapt plans as needed
- A good grasp of common sense knowledge and skills, e.g. cause and effect, defeasible reasoning, understanding of human feelings and values, ...
- Continual learning with models of the past, present and future
- Replacing prompt engineering by learning from the kind of responses most people prefer
- Reflective cognition utilising models of the agent's goals and performance in carrying them out, and likewise those of others (theory of mind)
- Able to explain itself in terms that we can easily appreciate, which may vary from one person to the next, e.g. needs to be age appropriate
- Adherence to the values we demand of them, e.g. we don’t want them to give racist, sexist and inflammatory responses, including instructions for making bombs, etc.

- We want AI agents to be unambiguously artificial agents, and not to be confused with humans.
- Smarter robots, self-driving cars, etc. with resilience to the unexpected
- As tools for boosting human creativity and effectiveness – better productivity for a more prosperous society if we share the benefits!
- As trusted personal agents that help us deal with a complicated world and look after our privacy, our finances and our health
- For stronger cybersecurity, and for countering disinformation, harmful content and conspiracy theories on social media
- AGI could one day win arguments with politicians and lawyers, leading to stronger democracies and better laws – doing so by in-depth access to knowledge, including which arguments will best convince people emotionally and intellectually*

* e.g. using classical rhetorical approaches, e.g. *ethos* (credibility), *pathos* (emotion), *logos* (logic), *kairos* (opportune) together with rhetorical questions

Future Neural Networks



Considerations

biggest questions are how to support episodic memory and continual learning

- To enable a mix of Type 1 and Type 2 processing along with a cognitive operating system
  - *how to manage time allocation for competing tasks akin to a mental operating system with feelings and drives?*
- Reflective cognition along with episodic memory to support situational awareness, including self-awareness and self-assessment in respect to execution of higher level goals
  - *It is easier to discuss sentience in the above sense than to discuss consciousness in general, which is harder to define, e.g. the so called "hard problem of consciousness" in respect to subjective experience (qualia)*

If all experience reduces to information processing with systems of neurons then perhaps qualia is a non-issue for artificial agents!

- The need for continual learning guided by reflective cognition
  - *inspired by human learning*
- How does the brain make memories?
  - *Episodic memory:* associative memory that can be used as a record akin to a personal diary, along with holding temporal relations that allow past experiences to be recalled in sequence*
  - *Encyclopaedic memory:* time-independent facts such as birds fly and dogs bark
- Episodic memory supports abductive reasoning, i.e. *what-if* thinking
  - *creating and updating plans, reasoning about cause and effect, inferring another agent's intent and state of mind*

* Retrieval of memories using a combination of what, where and when cues. Episodic memories are consolidated in the neocortex after initial modelling in the hippocampus. See: The Episodic Memory System: Neurocircuitry and Disorders (2010)

Lowering the Hurdles for Researchers

- LLMs with billions of parameters are very expensive to train
- Prohibitive for many researchers
- This is a barrier for work on innovative new network architectures
- A solution is to use smaller datasets and fewer parameters†
- Chosen to support research aims
  - Continual learning
  - Episodic memory
  - Reflective cognition

† See Kaggle report: Mini-giants: "small" language models (2023)

- Machine generated datasets
  - From LLMs, e.g. Microsoft's Tiny Stories*
  - From Knowledge Graphs using stochastic rules
  - Plus hand-crafted examples
- Different ways to learn
  - Observation, Instruction, Experience
- Evaluate different designs and select best for scaling up
- Small AI models are well suited for execution at the edge

* TinyStories: How Small Can Language Models Be and Still Speak Coherent English?, April 2023

Continual Learning

- Generative AI suffers from catastrophic task interference
  - Learning a new task dramatically degrades competence on previously learned tasks
  - Limited workarounds for transfer learning, which is also referred to as *fine tuning*
- Some potential solutions† include:
  - Weight regularisation
  - Sparse network connections
  - Lateral inhibition to free up neurons
  - Self-assembling neural networks*
  - Allocating tasks to neural modules akin to cortical regions with specialised roles
  - Meta-learning: learning to learn
- Giving AI agents dynamic access to models of the past, present and future – *aka* **episodic memory**

† Wang et al. survey of continual learning (2023) and Hospedales et al. Meta-learning in neural networks (2022)

- Memory for different time scales
  - Long term memory – neo*cortex*
  - Short term memory – *hippocampus*
  - Working memory – *activation levels*
- Perception related memory

  Baddeley and Hitch (1974, 1986)
  - Phonological loop – 1 to 2 seconds
  - Visual sketchpad – under one second
- Situational Awareness
  - Need for detailed short term memory
- Learning patterns across episodes
  - Need to avoid undue emphasis on most recent event vis a vis older events
  - Analogous to difference between the hippocampus and the neocortex

* Combining genetic algorithms with dynamic connections at run-time to mimic synaptic plasticity in vertebrate brains

Research Topics for Exploration

- Humans learn from few examples, but LLMs learn from vast corpora
- So called single-shot learning only applies to the prompt and doesn't update the model
- Can we mimic the roles of the Hippocampus and Neocortex?
- Separate models for short and long term memories
- Blend respective predictions in a manner akin to transfer learning

- The input and output of LLMs are tensors of shape *(batch size, sequence length, vocab size)**
- This assumes fixed vocabulary size with tokens as words, characters or some intermediary
- An alternative would be to use a vector database that maps encoded tokens to words or character sequences
- Support for learning on the fly

* Within the LLM, this is compressed to (batch size, sequence length, model width). For the output this is linearly transformed back to (batch size, sequence size, vocab size), and normalised into a probability distribution for stochastic selection of the best token.

Combining Feedforward with Feedbackward†

Feedback pathways are more numerous than feedforward pathways (Markov et al., 2014)

- Latent semantics, in the form of the activation levels of artificial neurons, can be seen as working memory, providing the context for word sense selection, prepositional attachment, attention, etc.
- Current LLMs use sequence lengths with many thousands of text tokens in purely feed-forward networks
  - Long range attention is expensive as pairwise attention scales quadratically
  - Feedback via progressively appending the generated tokens to the prompt
- Instead limit the encoder/decoder sequence length, and use feedbackward connections from latent semantics to lower layers
  - Mimicking human language processing

† Herzog, Tetzlaff & Wörgötter (2020): Neural networks in the brain are dominated by sometimes more than 60% feedback connections, which most often have small synaptic weights. Modern deep neural networks employ sometimes more than 100 hierarchical layers between input and output, whereas vertebrate brains achieve high levels of performance using a much shallower hierarchy. This may well be largely due to massive recurrent and feedback connections, which are dominant constituents of cortical connectivity.

- What kind of feedback* and why?
  - **Retained**: state held over from previous step, akin to RNN and LSTM
    - Key to sequential cognition (Type 2)
  - **Continuous**: as dynamic feedback
    - Key to language processing (Type 1)
- Plenty of Design Choices to Study
  - Is feedback implemented as multi-layer connections or as sequence of layer by layer transformations?
  - Transformers as integral to feedback?
  - Ensuring strong attractors for quick stabilisation during Type 1 processing, akin to collapse of a quantum superposition of states
  - Implications for deep learning?
- Heterogeneous neural network architectures
  - Featuring different kinds of neurons for different functional roles, e.g. short vs long term memory, and semantic vs spatial memory

* *see also:* Microsoft's RetNet (2023): Retentive Network: A successor to transformer for large language models; Hasani et al. (2022): Closed-form continuous-time neural networks

Human Inspired Language Models

- Human language processing is sequential, hierarchical and predictive
- 1 to 2 second capacity for the phonological buffer (Baddeley & Hitch, 1974)
- Implemented as small sliding window over text tokens
- Long range attention replaced by attention to latent semantics
  - Model producers and consumers to reduce cost of self attention
- Modules for words that understand their morphology
  - Associate words with their lexical features
  - Can be invoked to generate the characters*

* Have the language model decide on next word to generate and avoid need to run entire language model on each character

- Semantics **retained** from previous processing step and provided as layer input in combination with output from previous layer
  - Cheaper to process than dynamic feedback
- Informed by mathematical models of knowledge representation in vector spaces, e.g.
  - Use of vectors to generate Toeplitz matrices for circular convolution
  - Adaptive softmax inspired by Zipf's law†
- Aim to reduce cost of pretraining
  - Linformer: self attention with linear complexity (2020) Wang et al.

† Optimise computation on most common words, e.g. 87% of documents are covered by just 20% of the vocabulary

Additional Neural Modules

Sequential Cognition

- Neural equivalent to chunk rule engine
- Feed forward network that operates on latent semantics
- Corresponds to production rules with conditions and actions
- Learned from examples of reasoning and step-wise tasks
  - Deep reinforcement learning
- Working memory corresponding to deepest layer in language model
  - Input using values retained from previous processing step
  - Output used to update those values

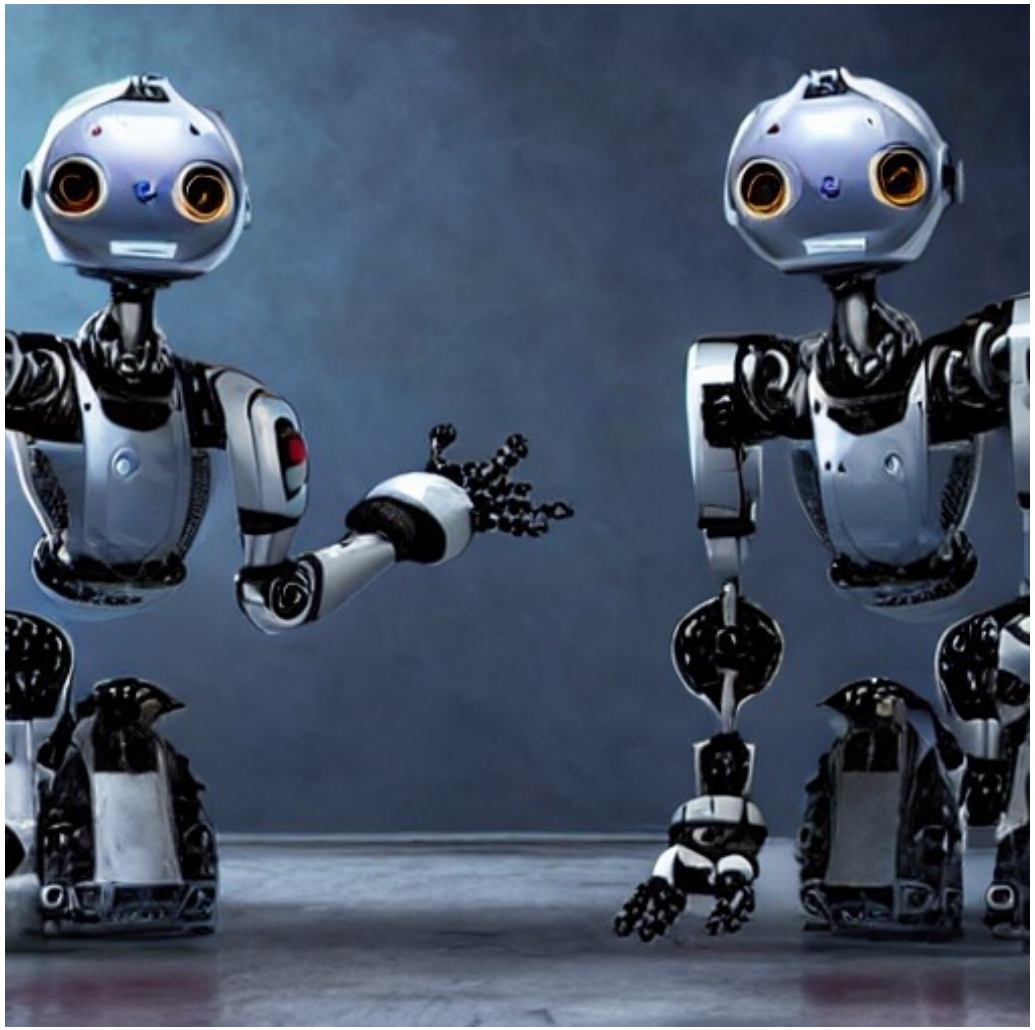
Episodic and Encyclopaedic Memory

- Acting as short and long term memory that supplements working memory
- Create, read, update and delete operations on vector database
- Additional operations as needed
- Mimics human forgetting curve
  - Stochastic recall based upon what proved most useful in past experience
- Episodic memory records salient details from snapshot of working memory
  - Based upon similarity metric to determine when to snapshot
  - Supplemented by relationships between episodes and process of memory consolidation*

* Memory consolidation was first referred to in the writings of Quintillian, a renowned Roman teacher of rhetoric. He noted the ”curious fact that the interval of a single night will greatly increase the strength of the memory,” and presented the possibility that” the power of recollection undergoes a process of ripening and maturing during the time which intervenes.”, *with thanks to Wikipedia*.

Semantic Interoperability

Knowing that we understand each other



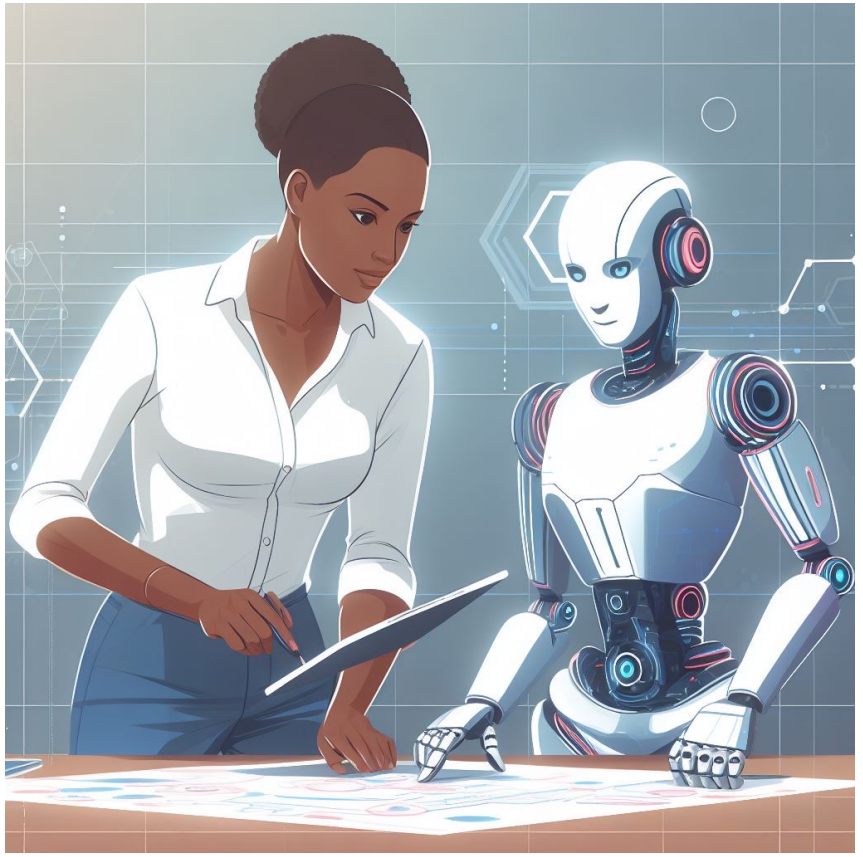
Generative AI lacks semantic consistency, as shown by the lack of support for the robot's body

Ensuring Mutual Understanding

- People keep written records when they don't want to rely on fallible memory
- The same applies to businesses
- Everyday language isn't good enough when we need to be sure of a mutual understanding
  - Business contract between a supplier and a consumer
    - Using standardised terms and legal language for contracts
- For technical exchanges we use structured data with agreed data models and semantics
- This relies on symbolic representations
- We will continue to need this as we make greater use of AI
- Knowledge Graphs as an evolution of databases
- Standardised vocabularies

Collaborative Knowledge Engineering

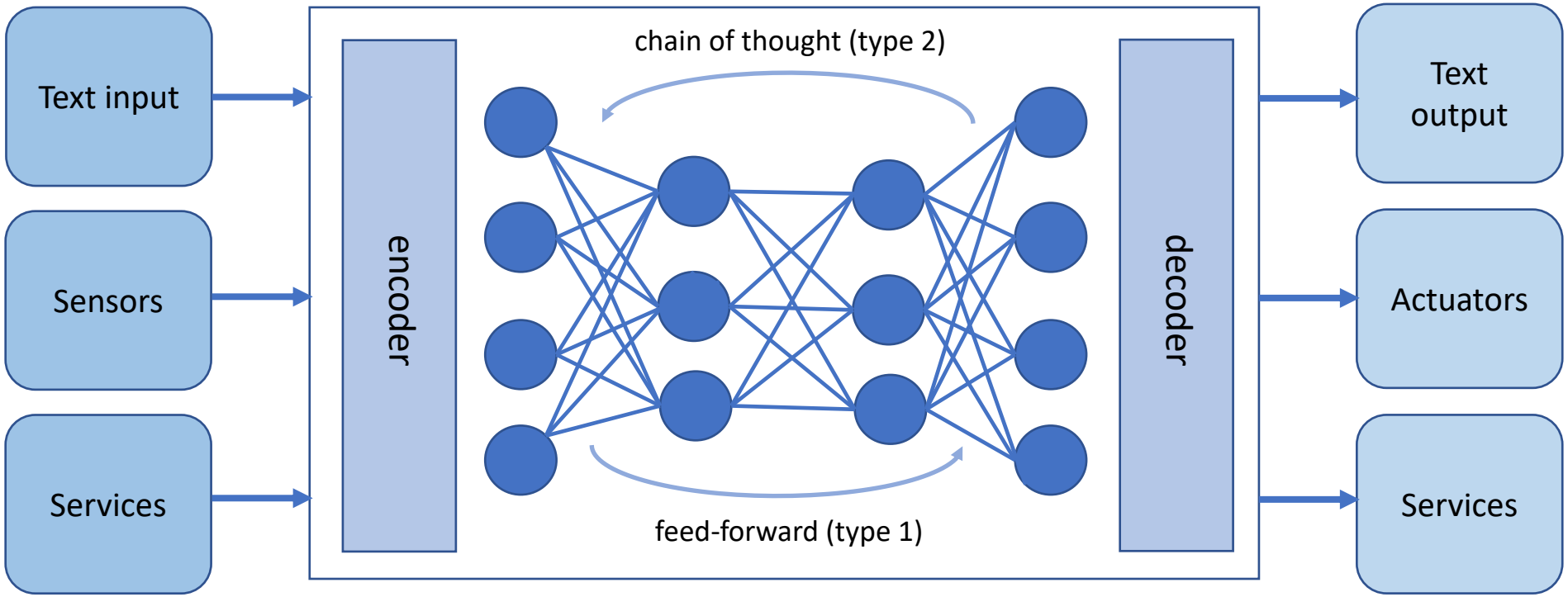
- Hand crafting knowledge graphs + rule sets is difficult and time consuming – this makes it hard to scale up
- Self-guided machine learning with neural networks is very much easier to scale up, but suffers from a lack of transparency
  - Knowledge is buried in the network parameters
- How can we use AI for collaborative knowledge engineering?
  - Human partner working together with an artificial agent
  - Agent operates on knowledge graphs + rule sets guided by human partner
  - Curating datasets, e.g. for new or updated use cases
  - Automated updates to rules as ontologies are revised
  - Versioning to support old and new applications



Note unsupported tablet floating in the air!

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. This can use comparatively smaller models that are distilled from larger models and fine-tuned 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.

Summary and Conclusions

- Expanding from logic to argumentation
- **Help wanted** with use cases and examples of good practices for argumentation
- Machine learning beats hand-crafted knowledge
- Ongoing role for symbolic AI for semantic interoperability
- Opportunities for work on human-like AGI
- **Help wanted** with use cases and training materials
- **Help wanted** with mathematical foundations for neural networks
- **Help wanted** with compute resources for training large models

Get in touch if you can help!

Discussion Topics

- **What will Human-like AI do for us?** It will be used • to make sense of the vast amount of information collected by the Internet of Things, • for cyber-physical control, e.g. self-driving cars, spaceships*, robots and machinery, • for human-machine collaboration as a means to boost creativity and productivity, • for education in support of human teachers, and • for personal agents that help us in an increasingly complex world.
- **What's the difference between Generative AI and Human-like AI?** Generative AI is statistical prediction based upon vast corpora of training materials. It suffers from weak semantic consistency, is easily distracted and prone to bias and hallucinations. Human-like AI, by contrast, seeks to support human-like reasoning, memory, learning, perception and actuation. This requires a much richer architecture compared to the transformer models used for large language models.

  Children often hazard a guess when answering questions, but can be taught to think things through step by step to avoid making glib mistakes. We can devise artificial cognitive agents that learn in a similar way through observation, instruction and experience. We can enable these agents to reason rigorously by carrying out checks on each step along the way.
- **Is there are role for small AI models?** Most definitely, when applied to specific roles and skills, and executed at the edge rather than the cloud.
- **What is the relationship between neural and symbolic based approaches?** Artificial neural networks have dramatically advanced our capabilities in respect to processing human languages. Linguists can relate their theories to computational models implemented in neural networks. Symbolic approaches have an ongoing role for semantic interoperability, to ensure communicating parties have a shared understanding using structured languages designed and maintained through human-machine collaboration. Symbolic approaches have been severely restricted by the emphasis on logic rather than argumentation. Neurosymbolic systems will have no such limitation, with the ability for machines to gather and apply vast amounts of qualitative metadata.
- **Does the Semantic Web need RDF?** The semantic web and RDF were predicated on the presumed need for machine interpretable annotations for websites as a basis for additional services. The most obvious example is smart web search based upon the schema.org vocabulary. However, websites can game this to achieve higher rankings in search results. Advances in AI make the need for such annotations less important as machines can now make sense of unstructured data in a variety of forms including text, images and audio. As such the Semantic Web needs a new vision that embraces AI and machine learning. Handcrafted RDF ontologies will soon look quixotic and a dusty relic of the past.

* e.g. the fictional onboard computer "Holly" in Red Dwarf

Questions and comments?

Contact: Dave Raggett <dsr@w3.org> W3C/ERCIM

European Research Consortium for Informatics and Mathematics
ERCIM

This work is supported by the European Union's Horizon research and innovation programme under grant agreement No. 101070487 for project Nephele on a synergetic meta-orchestration framework for the next generation IoT compute continuum.