Human-Like AI

Research directions for Webs of Cognitive agents that mimic human perception, memory, reasoning, learning and action

Dave Raggett, February 2023

With an emphasis on *research* rather than developing applications

Why is Human-Like AI Interesting?

Agents on the web refers to software programs that perform tasks on the internet. Examples of agents include web crawlers, which search the internet for information, and chatbots, which interact with users through text-based interfaces. Other types of agents include email clients, instant messaging clients, and even mobile apps that interact with web-based services (as written by ChatGPT)

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 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* internet of things!

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

What and Why

Boosting human productivity for increased prosperity for all as human populations age and shrink!

- **What is human-like AI?**
  - *Human-like AI, also known as artificial general intelligence (AGI), refers to the concept of creating a computer program or machine that has the ability to understand or learn any intellectual task that a human being can.*
  - *This type of AI would be able to perform a wide range of tasks and have the ability to adapt and learn new skills, much like a human being.*
  - *The goal of creating human-like AI is to build machines that can think and act like humans, with the ability to reason, make decisions, and solve problems in a human-like way.*
  - *However, achieving true human-like AI is still a subject of debate and ongoing research in the field of artificial intelligence.*

- **What is plausible reasoning?**
  - *Plausible reasoning is a type of reasoning that is based on what is possible or likely, rather than on certain knowledge or strict logical deduction.*
  - *This type of reasoning is often used in everyday life, as well as in fields such as law and medicine, where information may be incomplete or uncertain.*
  - *In AI, plausible reasoning can be used to make inferences or predictions based on incomplete or uncertain data, which can be useful in tasks such as natural language processing and decision-making.*

- **Traditional logic is not up to the challenge**
  - Everyday knowledge that is uncertain, imprecise, incomplete and inconsistent

Italicized text is generated by an AI – OpenAI's ChatGPT – showing that today's AI can work effectively with the semantics of natural language

Today's AI

- Dominated by artificial neural networks (ANNs) with billions of parameters, and trained on vast datasets, for example
  - ChatGPT – akin to a conversational version of Wikipedia
  - Minerva – undergraduate science problems
  - Stable Diffusion – image generation
  - Google Translate – natural language translation
- Larger models are qualitatively better, and can be further trained to meet specific needs

- Weaknesses
  - Knowledge is limited to original training datasets*
  - Limited to single tasks by catastrophic interference
  - Lacks introspection and has limited powers of reasoning
  - Relies on huge computing resources – which are very expensive and have a large carbon footprint
  - Dominated by a few wealthy organisations
- How can we design and train smaller models better suited to everyday needs?

* subject to bias reflecting limitations of training datasets

Plausible Reasoning with 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.

ChatGPT displays competence with natural language and plausible reasoning
Can this be duplicated on a smaller scale for lower computational cost?

Further examples

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.

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

Human-like Cognitive Agents

for human-machine cooperative tasks

- **Today**, we can chat with large language models, e.g. ChatGPT
  - Effective at a broad range of reasoning, but weak in respect to handling of trust and provenance
- But the long-term knowledge is static and limited to the set of resources used in training, e.g. a snap-shot of the Web
- Short term knowledge is limited to working memory – activation vectors for each layer in the network – and specific to the current dialogue with a particular user
- Large language models can be designed to support chain of thought reasoning, loosely corresponding to a mix of System 1 & 2
  - Parallel reasoning within each generative step
  - Sequential steps forming the chain of thought

- **What we want** are agents with continuous learning, and support for multiple tasks
  - Human-like memory to overcome catastrophic interference between new and old tasks
  - Human-like learning including co-occurrence patterns, taxonomic abstractions, skill compilation, and emotional intelligence
  - Richer kinds of reasoning, e.g. explanations and plans using causal, spatial, temporal knowledge etc.
  - Metacognition for flexible problem solving
  - Cyber-physical real-time simulation and control
- Support for short and long term memory specific to dialogues with a particular human
  - Need to reflect personal privacy, and confidentiality in respect to what cognitive agents learns from a given person, akin to attorney-client confidentiality
  - Generalises to groups of people, e.g. members of a business department playing a given organisational role requiring compartmentalising of reasoning to respect client confidentiality

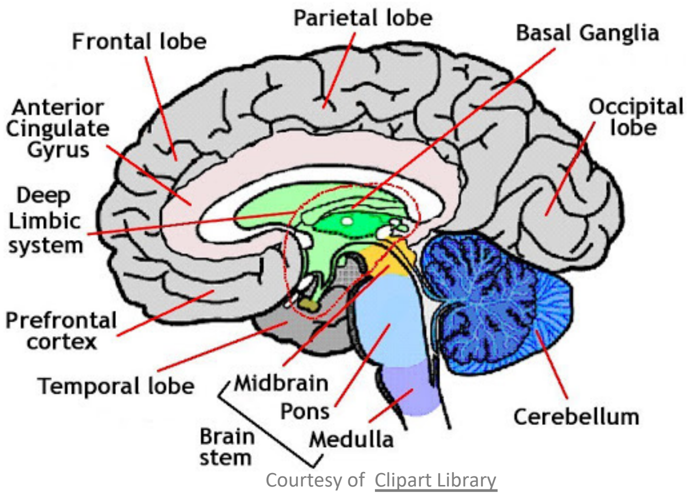
Communicating Agents*

High level approach for decision making in distributed systems

- Consider a group of people with different roles and collaborating on a shared task
- Each person decides on what actions they will take, based upon their physical senses and what they individually know
- People communicate with each other, sending and receiving messages
- People may enter or leave the group, as well as switching roles, as and when needed
- Knowledge fusion is the process of ensuring that each person has the up-to-date situational awareness needed to make effective decisions appropriate to their roles

* Related to the concept of swarm computing and ideas for hive minds with distributed agents with differing capabilities

- Now imagine replacing some or all of these people by software systems – cognitive agents on devices with differing capabilities
- Device sensors replace human senses, and we need to deal with heterogeneous communication technologies, data formats and protocols
- Knowledge graphs provide a representation that hides all of that complexity
- Perception involves progressively higher level interpretations of sensor data and messages from other agents
- Cognition is the process of deciding on what actions to take
- Actions include messaging other agents as well as real-time control over actuators

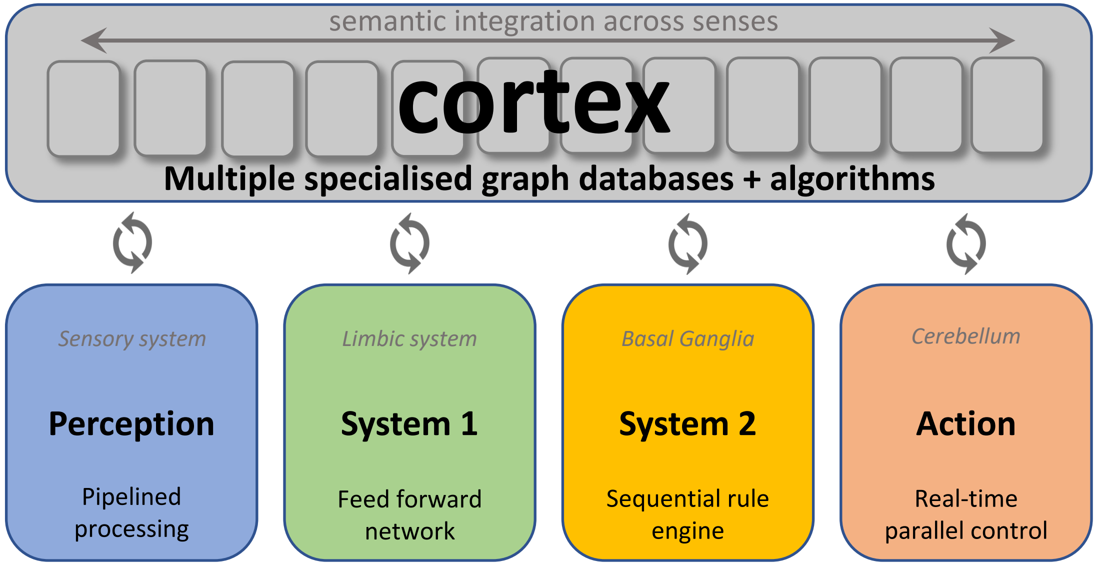


Anterior temporal lobe as hub for integration across senses

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.

Mimicking Human Memory

- Humans can learn multiple tasks with minimal interference
  - Ebbinghaus forgetting curve
  - Spacing effect
  - Stochastic recall
  - Semantic priming*
  - Contextualization
  - Emotional connotations
- Today's artificial neural networks are subject to catastrophic interference
  - Training a new task destroys performance on old tasks
  - Impedes continuous learning and extensions for general problem solving

- Some ideas include
  - Richer models of neural dendrites
  - Reduced plasticity for connections that are important for tasks
  - Inhibitory connections that free up neurons for other tasks
  - Rehearsal using generative replay to reinforce old tasks – akin to REM phase of human sleep†
  - Using different networks for short and long term memories
    - Mimicking Hippocampus, Cortex and Amygdala in mammalian brains‡
      - Hippocampus for short term memory
      - Cortex for long term memory
      - Amygdala links short and long term memories to emotional connotations

* through spreading activation from related concepts

† responsible for memory consolidation during sleep

‡ see MIT study of memory formation (2017)

Further Considerations

- Large artificial neural networks are clearly able to model semantics for natural language and for images
  - Images as hierarchical compositions
  - But weaker on causal understanding
- The neural activation levels provide a distributed form of working memory
  - Opaque representation of latent semantics across multiple network layers
- Corresponds to System 1 cognition
  - intuitive, fast, sometimes wrong, apparently effortless, and opaque in respect to how conclusions were reached
- New work is needed to enable operations on latent semantics
  - e.g. insert, remove, access, update and count parts of image compositions
- This is needed to improve sequential reasoning for System 2 cognition*
  - Slow, deliberative, analytic, open to introspection, effortful, supplementing System 1 cognition
- With further work needed on
  - Continuous learning
  - Metacognition
  - Learning how to reason

* Large language models can emulate System 2 with sequential chain of thought responses

Other Challenges

- Neural networks are great at learning hierarchies of conditional statistics and associated functions
- But cardinality is more challenging*
  - e.g. number of fingers on each hand
- What kind of network architecture would make it easier to learn and apply cardinality constraints?
  - Evolving beyond TRANSFORMERS
- Likewise for causality constraints
  - How to enable richer understanding that goes beyond modelling correlations

"photo of a hand on a white sheet" by Stable Diffusion

* In RDF you can use owl:qualifiedCardinality and sh:minCount, etc.

Incorporating Domain Knowledge

- How to combine deep learning with explicit domain knowledge?
- One approach uses a set of questions and worked answers, where latter are given as a sequence of derivation steps, e.g. as for Google AI's Minerva
- System predicts next step based upon current step and context at different levels of abstraction
- A complementary approach maps domain knowledge expressed in terms of properties, relationships and implications to latent semantic space
- This mapping is trained in parallel with problem solving involving sequences of transformations of working memory, i.e. current neural activation vectors
- What about dynamic knowledge?

Systematic Generalisability

- According to Greff et al. (2020)* existing neural networks fall short of human-level generalization, requiring large amounts of data, struggle with transfer to novel tasks, and are fragile under distributional shift
- Under the right conditions, they have shown a remarkable capacity for learning and modelling complex statistical structure in real-world data
- One explanation for this discrepancy is that neural networks mostly learn about surface statistics rather than relationships between the underlying concepts, which prevents them from generalizing systematically
- Current neural networks appear to lack the means to bind information through variables
  - But does this still apply to recent LLMs?
- This relates to the use of symbols as a basis for relating and manipulating mental entities
- Neural networks ground symbols in input data
- Distributed object representations, e.g. using circular convolution over an object's properties
- But how to persist object identities over time despite changes to the object's properties?
- Different kinds of properties: instance, temporal, spatial, category
- Relationships, e.g. X left-of Y, X part-of Y
- Implications as if-then rules with variables
- Multi-stable segregation and coherence for understanding images and text
  - Potential role of dynamic oscillatory models?

* On the Binding Problem in Artificial Neural Networks

Enabling Introspection

- Create a smaller version of Stable Diffusion for a very limited range of images
  - Millions rather than billions of parameters
  - Designed for training on everyday computers
  - Designed to handle collections and quantities*
- Algorithmic generation of prompts and images for the training data
  - Along with richer descriptions of image components
- Integrate with a neural network designed for operating on latent image semantics
  - Train it to apply operations such as: insert, remove, update, count, select, iterate and describe
  - Including selection by pointing
- Apply language model to support natural language operations on latent semantics
  - Using an existing model, e.g. DistilBERT *or* one trained on algorithmically generated dataset

* e.g. "two red discs and three blue squares", which is poorly handled by Stable Diffusion

Latent Semantics and Working Memory

- Multi-layer networks for large language models*
  - Lower to middle layers capture information about syntactic structure
  - Upper layers are more important in respect to predicting following text
  - How to enrich taxonomic knowledge?
- Want to mimic humans in respect to language processing
  - GPT, BERT, BLOOM, etc. all process words in parallel, not incrementally
  - Incremental word by word model relies on working memory for past syntactic and semantic context
  - How then to support attention?
  - Insights from shift-reduce parsing
    - Preparedness rather than prediction
- Similarities with natural language translation using sequence to sequence mapping
  - Adding further layers on top of the layers used for the language model
- How does working memory encode temporal information, e.g. previous utterances?
  - How large is working memory?
  - Continuous learning as a way to remember what's useful without overloading working memory
    - Short term versus long term memory
    - Episodic and context based memory
- Using visual saccades for modelling attention in image understanding?
  - Complementing transformers
    - Sequential processing models

* Interpretable semantic representations from neural language models and computer vision, Steven Derby, 2022

Sequential Reasoning

See also "chain of thought" operation of large language models

- Train neural network to store and recall production rules*
  - Operate on working memory, i.e. directly on latent semantics
- Determining which rules apply to working memory based on rule conditions
- Selecting the "best" rule
  - Stochastic and biased by experience
- Applying rule's actions to update working memory, long term memory, invoke external actions

- Sequential reasoning for
  - Metacognition – reasoning about reasoning and problem solving
  - Plausible reasoning and lines of argumentation
  - Reasoning about explanations and plans, including theory of mind
- Continuous learning
  - Co-occurrence patterns
  - Taxonomic learning
  - Skill compilation
  - *Lessons, observations, experience**
- External actions
  - Delegated to separate system capable of real-time control loops

* Mimicking ACT-R and Cognitive AI CG's chunks & rules

* Children are very observant and learn from playing

Causal Reasoning and Physical Actions

How does the cortico-cerebellar circuit support explaining, learning, imagining, planning and performing physical actions?

- Understanding cause and effect is key to explaining what's happening, and to planning how to achieve some set of goals.
- Causal knowledge can be learned by constructing models that provide plausible explanations, and which are effective when applied to executing plans.
- Causal reasoning further requires a conceptual treatment of time, including relative times, and open and closed intervals.
- If X occurs before Y, then perhaps X is involved in causing Y, or perhaps something else is responsible for both X and Y, or even that X and Y occur randomly, and no causal connection is involved.
- Causal relationships need to have a plausible basis, i.e. an understood means of causation, e.g. pulling or pushing as a means to make something move.
- A further consideration is motivation and intent when reasoning about sentient agents. Why did this agent take that action? This is part of the theory of mind which covers the beliefs and actions of others.
- Physical actions involve the complex real-time coordination of many muscles guided by perception, including vision, touch and proprioception.
- We can imagine taking an action without actually performing it. This also occurs when we see someone else performing an action, e.g. smiling, laughing, singing, jumping and dancing. This allows us to mimic the actions of others.
- This is a form of causal reasoning that models the physical actions an agent would need to take, including smooth functions over time, repetitions and rhythms – imagining as simulation.

Importance of gathering different kinds of use cases for causal knowledge and reasoning

Comparison

Symbolic Approaches

- Symbolic approaches based upon graphs, e.g. RDF with labelled directed edges
- Challenging to scale up due to need for manual knowledge engineering
- Lack robustness when faced with things not considered by the developers
- Traditional logic is not that useful in practice, so reasoning is usually hand-crafted in application code, making it expensive to adapt to changes
- Easy to support quantitative constraints
- Require extension to support imperfect and imprecise knowledge*

Artificial Neural Networks

- Scale well using large datasets, but requires more powerful computers
- Automates knowledge engineering
- But, likewise, lack robustness on out of distribution data, and current solutions need much more data than humans
- Existing networks are weak in respect to reasoning, System 2 cognition, and lack introspection and continuous learning
- Hard to support quantitative constraints
- Good at handling context sensitive and imprecise concepts

* see plausible knowledge notation and demo

Where Next for Symbolic Approaches?

- Continue working on plausible reasoning for representing and reasoning with knowledge graphs
  - Combining symbolic knowledge, qualitative metadata and human-like argumentation
- See how easy it is to mimic intuitive reasoning, e.g. for natural language understanding
  - Natural language is very much easier than working with RDF/Turtle and OWL
- Extension for different kinds of reasoning, continuous learning and metacognition
  - Causal knowledge and reasoning
- Exploring strengths and weaknesses relative to artificial neural networks
  - Driving progress in both fields
- Importance of semantic interoperability for scalable deployment
  - Explicit knowledge and context sensitive reasoning
  - Heterogeneous vocabularies reflecting differing requirements
- Evolutionary replacement for RDF
  - Plausible Knowledge Notation is higher level and more flexible than RDF/Turtle
  - Natural language for everyday users

But isn’t it better to focus on extending neural networks, given how much better they are now at representing and reasoning with human knowledge

What comes after the Semantic Web?

- **Symbolic plus Sub-symbolic knowledge**
- Extending neural networks with explicit domain knowledge
- Continuous learning and improved reasoning
- Natural language interaction

Higher Level Notations for Symbolic Knowledge

- Simplifying level of expertise needed to develop knowledge-based applications
- What comes after spreadsheets – how about ***Knowledge Sheets***?
- Easy to use, and integrated with enterprise knowledge graphs
- Better aligned to semantics of natural language and everyday knowledge, as key to cooperative human-machine interaction
- Representing and reasoning with imperfect and imprecise knowledge, i.e. everyday knowledge in an imperfect world
- **Plausible reasoning** (Alan Colins), **Chunks & Rules** (John Anderson)

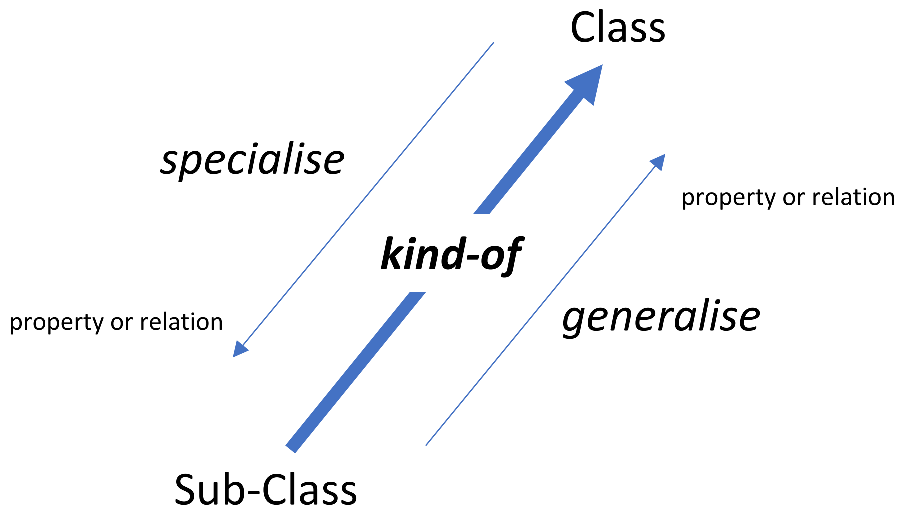
Plausible Knowledge Notation

Simpler, more powerful, and higher level than RDF

See Web-based demonstrator

Plausible Inferences using Prior Knowledge

- Inferring likely properties and relations across other relations



- Expected certainty influenced by qualitative metadata
  - e.g. typicality, similarity, strength, …

- 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
- Scalar ranges
  - fuzzy terms, e.g. cold, warm and hot
  - fuzzy modifiers, e.g. *very* old
- Multiple lines of argument for and against the premise in question

eBNF Grammar for Plausible Knowledge Notation (PKN)

statements, queries, analogies

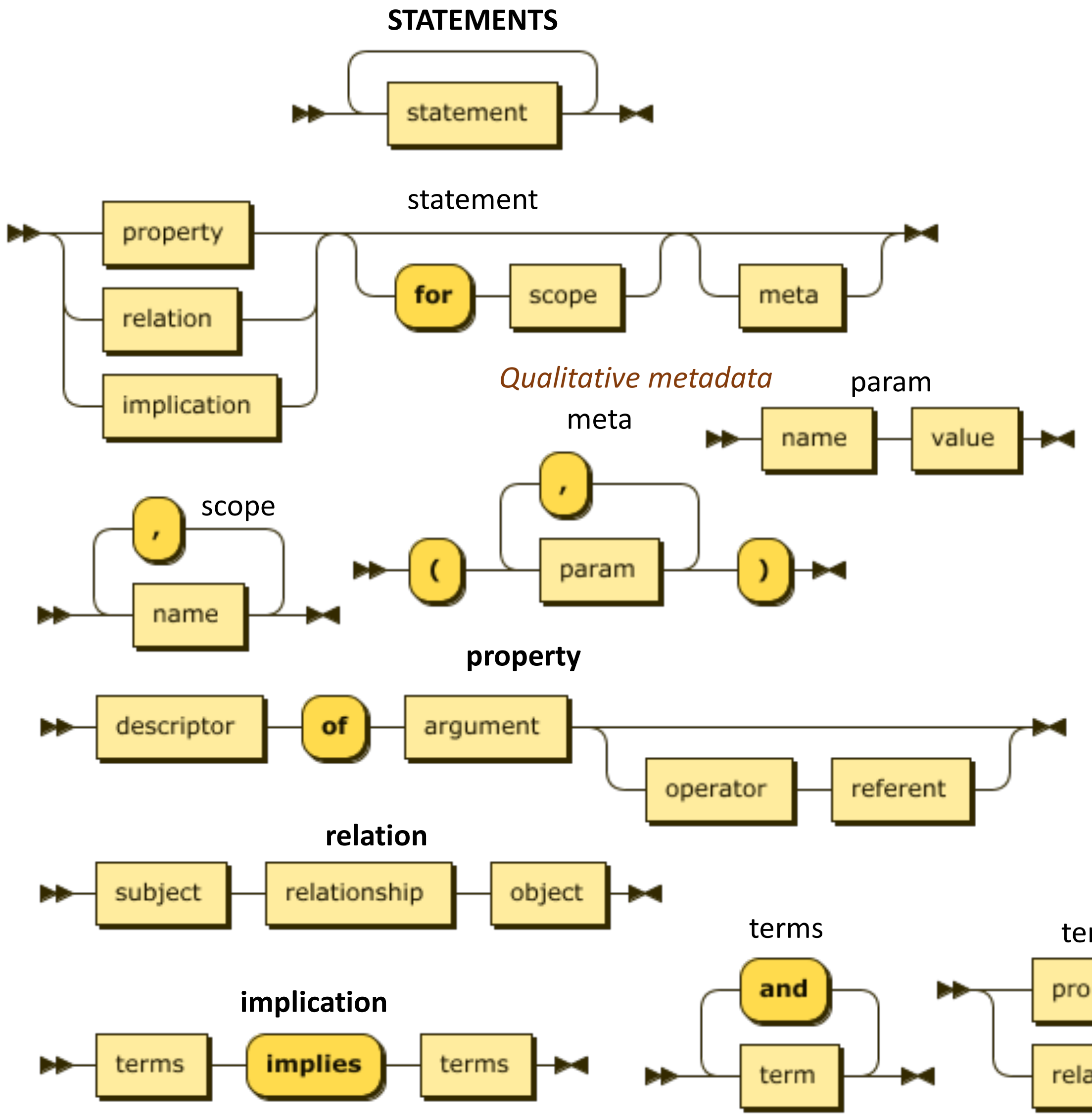
```
STATEMENTS ::= statement+

statement ::= (property | relation | implication) ("for" scope)? meta?
   scope ::= name (',' name)*
   meta ::= '(' param (',' param)* ')'
   param ::= name value

property ::= descriptor "of" argument (operator referent)?
   descriptor ::= name | variable
   argument ::= name | variable
   operator ::= "includes" | "excludes " | "is" | name
   value ::= name | number
   referent ::= (value ("," value)*) | variable

relation ::= subject relationship object
   subject ::= name | variable
   relationship ::= name | variable
   object ::= name | variable

implication ::= terms "implies" terms
   terms ::= term ("and" term)*
   term ::= property | relation
```

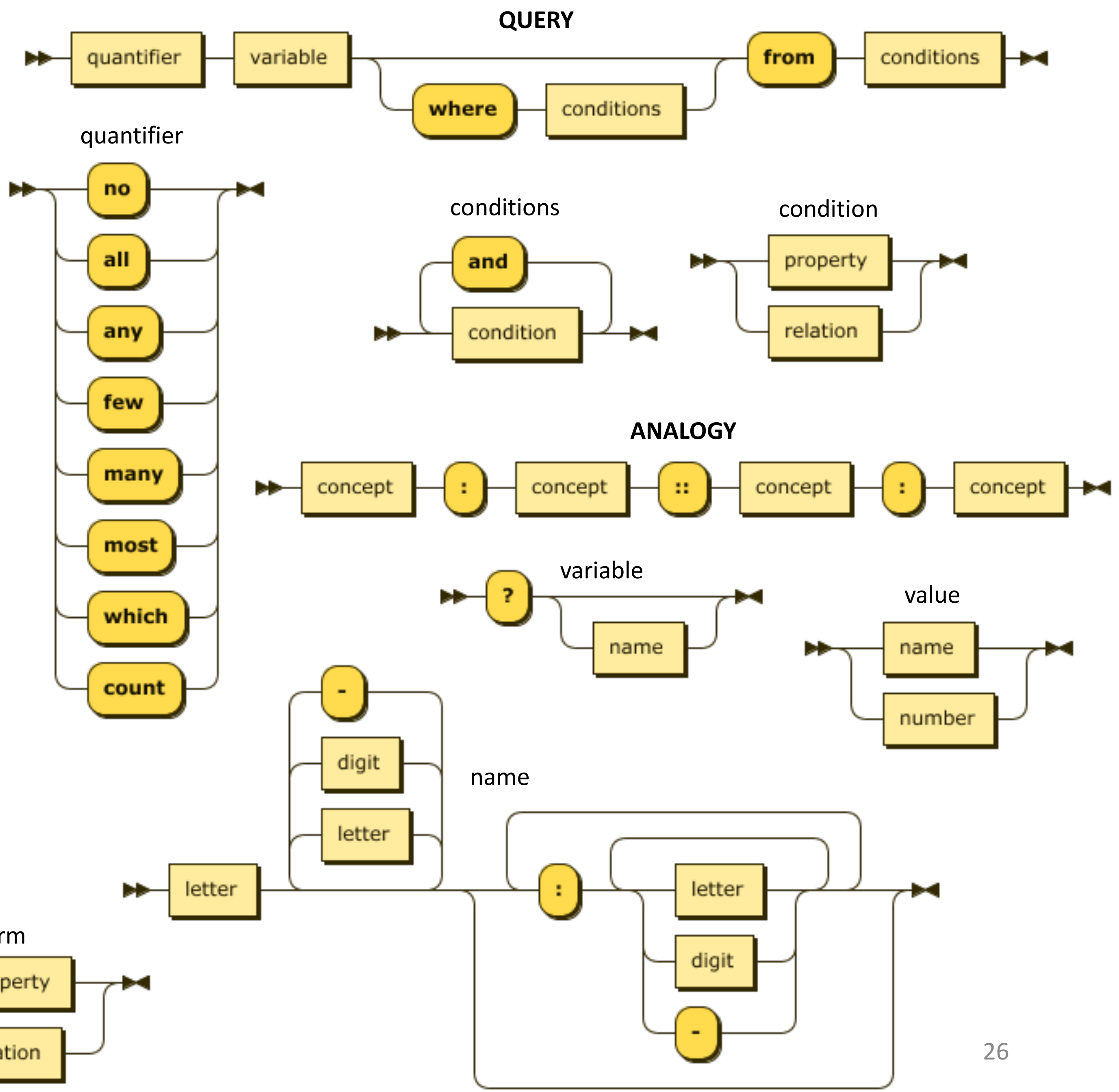
```
QUERY ::= quantifier variable ("where" conditions)? "from" conditions
   quantifier ::= "no" | "all" | "any" | "few" |
                  "many" | "most” | "which" | "count”
   conditions ::= condition ("and" condition)*
   condition ::= property | relation

ANALOGY ::= concept ":" concept "::" concept ":" concept
   concept ::= name | variable

variable ::= "?" name?
name ::= letter (letter | digit | "-")* (":" (letter | digit | "-")+)*
number ::= ("-" | "+")? digit+ ("." digit+)?

letter ::=
   "A" | "B" | "C" | "D" | "E" | "F" | "G" | "H" | "I" |
   "J" | "K" | "L" | "M" | "N” | "O" | "P" | "Q" | "R" |
   "S" | "T" | "U" | "V" | "W" | "X" | "Y" | "Z" |
   "a" | "b" | "c" | "d" | "e" | "f" | "g" | "h" | "i" |
   "j" | "k" | "l" | "m" | "n" | "o" | "p" | "q" | "r" |
   "s" | "t" | "u" | "v" | "w" | "x" | "y" | "z"

digit ::= "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"
```


variables are scoped to the implication or query in which they occur

PKN grammar

statements, queries, analogies





PKN Examples from Web 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
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

Enriched semantics and easier to use notation relative to RDF/turtle

Model vs Data Driven Natural Language

Natural language interaction is much easier for most people*

- What is needed to support collaborative dialogues between human users and cognitive agents?
- Grice's maxims for cooperative dialogues
- Lexical knowledge for words and their meanings along with their co-occurrence statistics
- Syntactic processing – e.g. shift-reduce parsing
- Semantic processing – constructing a coherent plausible semantic model of a natural language utterance based upon prior knowledge
- Humans can understand text and speech very quickly using intuitive and opaque processing (System 1) prior to slower deliberative reasoning (System 2)
- Large language models learn how to do that from self-supervised training against huge corpora of text documents
- Can we duplicate that for symbolic systems as manual knowledge engineering is inefficient and slow?

Some examples involving common sense knowledge

P: John took an umbrella with him

Q: Why did he take an umbrella?

A: Because he thought it would rain

or because it was raining or the forecast was for rain.

P: Sue put the kettle on

Q: Why did she put the kettle on?

A: Because she wanted to make some tea

P: John opened the bottle and poured the wine.

Q: Why did he open the bottle?

A: In order to pour the wine

Q: What happened to the wine?

A: It was poured into the glasses

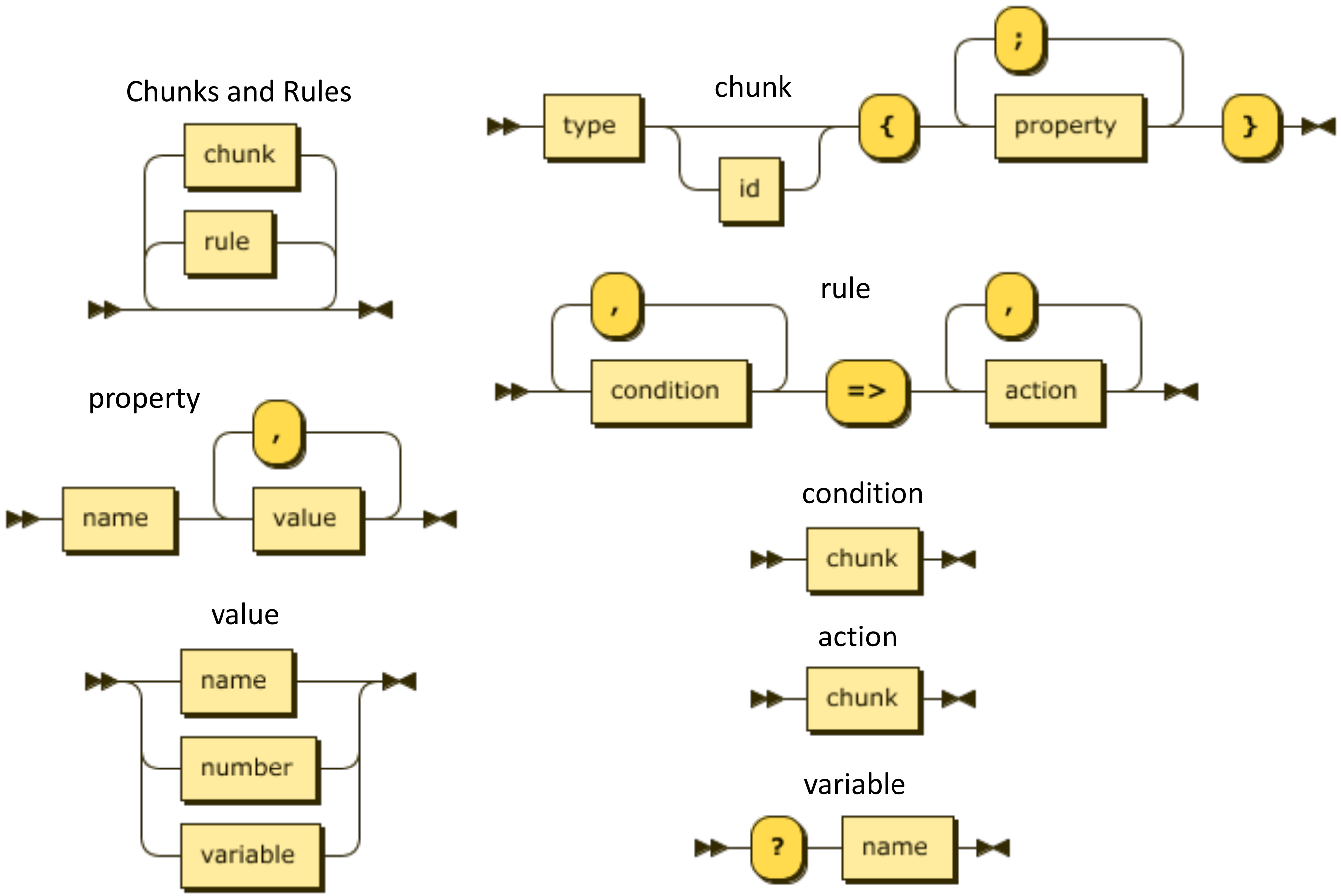
* As compared to programming languages

Chunks and Rules

A model of sequential cognition

Chunks and Rules

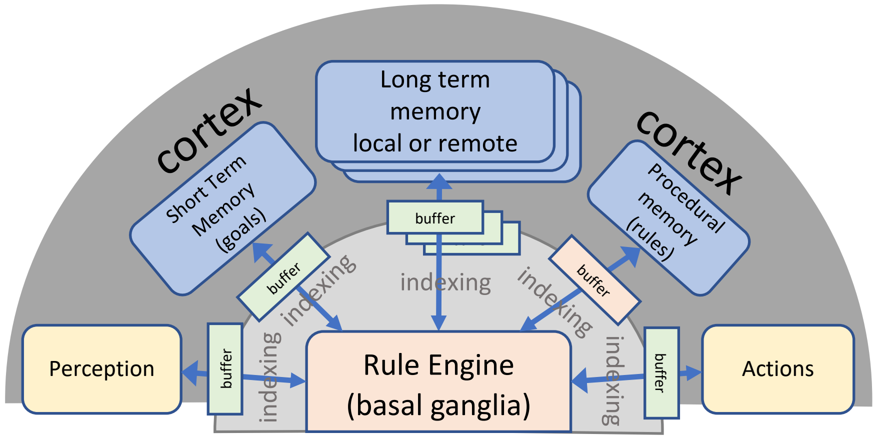
web-based demos for smart homes and factories



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