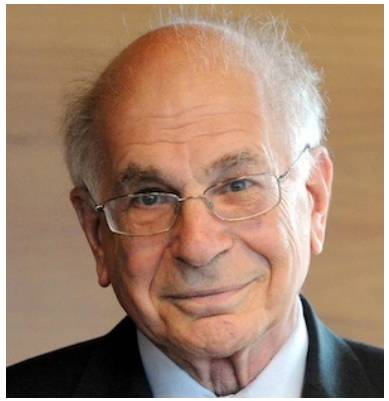
Discussion Topics

workshop on imperfect knowledge

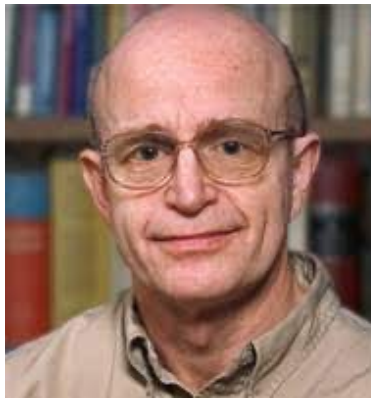
03 May 2022, as part of KGC-2022

Some Pioneers

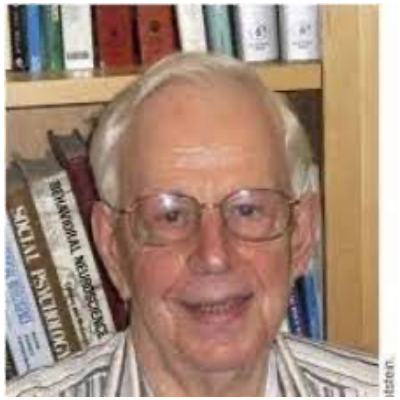
We owe them a huge debt for their hard won insights



Daniel Kahneman, Nobel prize winning psychologist who studied System 1 & 2 thinking, along with cognitive biases in “Thinking Fast and Slow”.



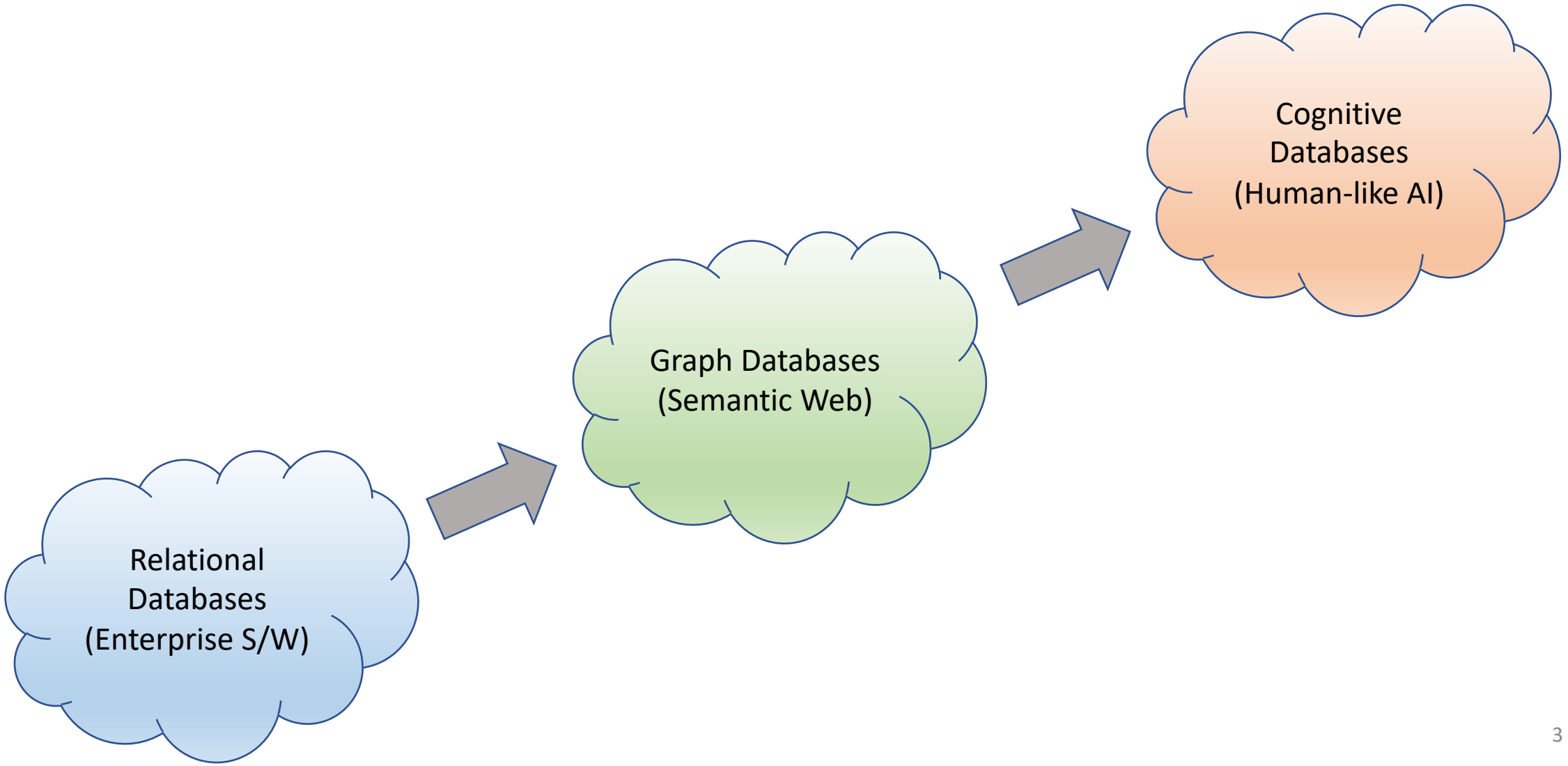
John R. Anderson, cognitive scientist renowned for his work on the ACT-R cognitive architecture for sequential cognition (System 2).



Philip Johnson-Laird, cognitive scientist renowned for his work on how humans reason in terms of mental models rather than using logic and statistics.

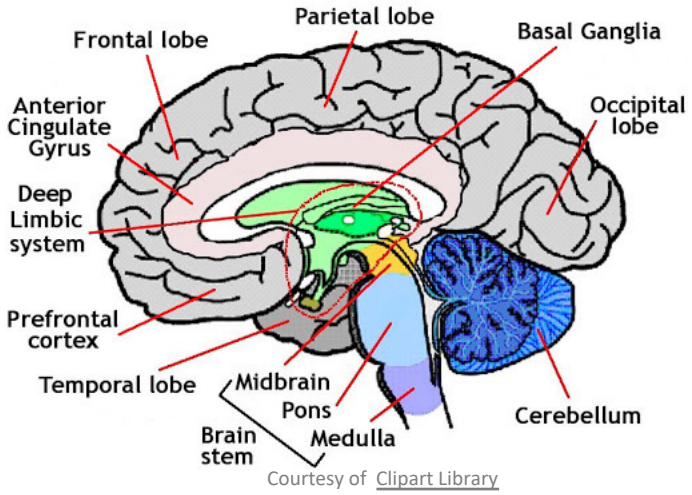
Allan Collins, cognitive scientist renowned for his work on plausible reasoning and intelligent tutoring systems.

Evolution in Action

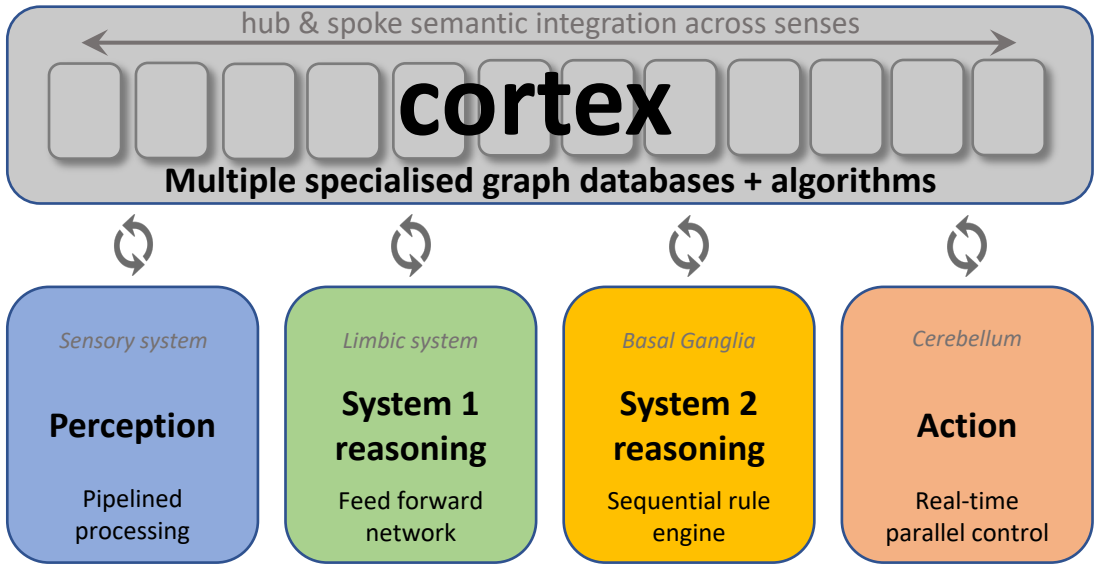


Cognitive Architecture for human-like AI

Basis for next generation graph database systems



Cognitive Architecture with multiple cognitive circuits loosely equivalent to shared blackboard



Semantic integration mimics Anterior Temporal Lobe's role as a hub for unimodal spokes

Memory is based on graph databases and associated graph algorithms. It combines symbolic graphs with sub-symbolic information, mimicking the human cortex, and defined at a conceptual level above that of RDF and Property Graphs (including NGSI-LD). Recall is stochastic reflecting prior knowledge and past experience. This involves activation boost/decay, spreading activation, the forgetting curve and spacing effect.

Perception interprets sensory data at progressively higher levels of abstraction, 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 is used to avoid missing closely spaced events.

System 1 is about cognitive control and prioritising what's important. The limbic system provides rapid automatic assessment of past, present and imagined situations without the delays incurred in deliberative thought. 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 actively guided by perception.

System 1 and 2

We can use human reasoning as inspiration for cognitive agents that think and learn like us, but don't lose concentration and are better at countering cognitive biases

System 1 (intuitive/emotional)

- Operates automatically and apparently effortlessly using parallel processing
- Closely related to human perception and associative memory
- Used to rapidly construct coherent mental models of past, present and imagined situations
  - Including natural language understanding/generation
- Subject to systematic cognitive biases
- Surprise detection leveraging continuous learning
- Syntagmatic vs Paradigmatic learning
- We are prone to see causal agency even in random fluctuations
- What kind of causal reasoning is directly supported by System 1?

System 2 (analytical thought)

- Sequential, deliberative and slow
- Requires effort and is easily distracted
- Used to verify suggestions from System 1, but is often lazy, accepting suggestions without checking them
- Implementable using chunk rules with highly constrained working memory
  - Suite of web-based demos
- How is plausible reasoning layered on top of System 1 and 2?
  - Reasoning that is partly automatic and partly under deliberate conscious control
- How is metacognition layered on System 2
- Role of emotions and drives for control

Deep Learning isn't sufficient to explain System 1

Characteristics of System 1

according to Daniel Kahneman

Plenty of clues as to how System 1 works

- Generates impressions, feelings and inclinations; when endorsed by System 2, these become beliefs, attitudes and intentions
- Operates automatically and quickly, with little or no effort, and no sense of voluntary control
- Can be programmed by System 2 to mobilise attention when a particular pattern is detected *(search)*
- Executes skilled responses and generates skilled intuitions after adequate training
- Creates a coherent pattern of ideas in associative memory
- Links a sense of cognitive ease to illusions of truth, pleasant feelings and reduced vigilance
- Distinguishes the surprising from the normal
- Infers and invents causes and intentions
- Neglects ambiguity and suppresses doubt
- Is biased to believe and confirm
- Exaggerates emotional consistency *(halo effect)*
- Focuses on existing evidence and ignores absent evidence *(what you see is all there is)*
- Generates a limited set of basic assessments
- Represents sets by norms and prototypes, but doesn't integrate
- Matches intensities across scales *(e.g. size to loudness)*
- Computes more than intended *(mental shotgun)*
- Sometimes substitutes an easier question for a difficult one *(heuristics)*
- Is more sensitive to changes than states *(prospect theory)*
- Over weights low probabilities
- Shows diminishing sensitivity to quantity *(psychophysics)*
- Responds more strongly to losses than to gains *(loss aversion)*
- Frames decision problems narrowly, in isolation from one another

Common Cognitive Biases

Further clues as to how System 1 works, and how to mitigate bias

- **Confirmation bias** - we pay more attention to information that confirms our existing beliefs
- **Hindsight bias** - we see events as more predictable than they actually are
- **Anchoring bias** - we pay too much attention to the first piece of information we hear
- **Misinformation effect** - our memory is often influenced by later events, and we tend to fill in gaps to flesh out a coherent account
- **Actor-observer bias** - we tend to attribute our actions to external factors and other people's actions to internal ones
- **False consensus effect** - we over estimate how much other people agree with us
- **Halo effect** - the tendency to rely on initial impressions, e.g. good-looking people are seen as smarter and more reliable
- **Self-serving bias** - we credit ourselves for successes, but blame others for our failures
- **Availability heuristic** - we over estimate probabilities based upon easily recallable memories
- **Optimism bias** - we over estimate the likelihood that good things will happen to us, whilst underestimating the likelihood of negative things

Plausible Reasoning vs Deductive Logic

Deductive Logic

- The basis for the Semantic Web
- Mathematically based with automatic proof procedures
- But doesn't scale well when applied to higher order logics
- Doesn't consider prior knowledge* thereby limiting what inferences are permitted
- Assumes Boolean truth values and doesn't support uncertainties
- Fails with inconsistent knowledge

Plausible Reasoning

- Human-like reasoning both for and against a given premise
- Heuristic approach to assessing different lines of arguments
- Takes prior knowledge into account for flexible inferencing
- Works with uncertain, incomplete and inconsistent knowledge
  - Relation to cognitive dissonance, and decisions to change goals or beliefs
- Works with higher order statements, e.g. theory of mind

* e.g. baseline statistics, typical examples etc.

Qualitative Metadata

Used to estimate certainty for each plausible inference
the algorithm combines multiple sources of evidence*

- *Typicality* in respect to other group members
  - e.g. robins are typical song birds
- *Similarity* to peers
  - e.g. similar climate
- *Strength* – conditional likelihood
  - e.g. strength of climate for determining which kinds of plants grow well
- *Frequency* – proportion of children with given property
  - e.g. most species of birds can fly
- *Dominance* – relative importance in a given group
  - e.g. size of a country's economy
- *Multiplicity* – number of items in a given range
  - e.g. how many different kinds of flowers grow in England

* This is not as simple as it sounds, and can be combined with baseline statistics!

Qualitative vs Quantitative metadata

Quantitative Metadata

- Based upon statistical theory, e.g. Bayesian inference for hypothesis H and evidence E

  $$P(H|E) = P(E|H)\,P(H) / P(E)$$

- The required statistics may not be easily available, e.g. likelihoods and prior probabilities
- How does Bayesian inference apply to different kinds of plausible inferences?
- How to combine qualitative and quantitative metadata to get the best of both techniques?
  - Starting with tracking baseline statistics

Qualitative Metadata

- Human reasoning isn't based upon logic or statistics, but rather on mental models involving typical examples and analogies
  - Baseline statistics are often ignored*
- Informal approach to computing certainty of a given inference
- Heuristic parameters which are applicable to different kinds of inferences
  - *typicality, similarity, frequency, dominance, multiplicity, strength*
  - Humans make plausible guesses for these based upon examples, and subject to a range of cognitive biases

* People pay more attention to individual examples and stereotypes as demonstrated in experiments by Kahneman and others

Knowledge Graphs, Syntagmatic and Paradigmatic Learning

- Plausible models of episodes
- Syntagmatic Learning deals with collocational statistics within each episode, e.g. nouns and verbs in the same utterance
- Paradigmatic Learning deals with taxonomic abstractions, e.g. pets as a super class for dogs and cats
- Children exhibit syntagmatic learning at an early age and gradually develop paradigmatic learning abilities

- How to replicate this with knowledge graphs?
  - Symbolic & subsymbolic information
- How to model episodes?
  - Separate graphs?
  - Imagined contexts?
- Taxonomic learning
  - Concepts, properties & relationships
  - Class hierarchies
- Potential advantages of neurosymbolic approaches

Causal Knowledge and Reasoning

- Different kinds of causal knowledge
  - Naïve physics, everyday plans, social reasoning and large scale medical trials
- Graph edges can be used to express causal connections and vertices to represent cause and effect
  - On/off as for light switches
  - Qualitative values, e.g. the gas setting for a kitchen hob: off, low, high
  - Quantitative values, e.g. 0 to 9
- Qualitative modelling of physical processes
  - Vertices as processes rather than states, and edges as transitions, e.g. when water starts to boil, or has boiled away
- Fuzzy reasoning with ensembles of states akin to quantum systems
  - Certainty: 70% hot, 30% cold
- Dealing with ambiguities
  - Ambiguities when seeking likely explanations through abduction from observed state, and when predicting likely behaviour
- Surprise detection
  - When observations and predictions don't match
  - Continuous learning as a background process
  - We only consciously remember the unexpected!
- Learning causal knowledge
  - Judea Pearl's three layers for causal learning
    1) observing temporal correlations, 2) trying things out, and 3) imagining different possibilities
  - Automatic update of subsymbolic metadata
  - Belief revision when required by evidence
- System 1 and 2
  - Automatic vs deliberate causal reasoning
  - What use cases can help clarify this?

Humans see causation and agency/intent pretty much everywhere they look!

Natural Language

- Important for flexible human – cognitive agent collaboration
- Large Language models, e.g. GPT-3, aren't effective in respect to reasoning
- We instead need to use plausible reasoning to map natural language utterances to graphs expressing the intended meaning
  - Along with being able to reason about explanations
- This is fast, automatic and coherent, hiding ambiguities in favour of most likely meaning
  - i.e. part of System 1 rather than System 2
- Word by word incremental processing
- Parsing is easy if you leave ambiguities to asynchronous processing of meaning
  - Word senses, prepositional attachment, referring expressions
- Language generation following Grice's maxims for cooperative dialogue
  - Quantity, quality, relation and manner

- Consider the following example

  John opened the bottle and poured the wine.
- Using plausible reasoning, we can infer that:
  - The above utterance very likely relates to a social occasion such as a dinner or a party where it is customary to drink alcohol
  - John is likely to be a host
  - He, as an agent with intent, caused the bottle to open and then caused the wine to flow from the bottle to the guest's glasses
  - Opening the bottle is a precondition for pouring its contents
  - The bottle contained wine at the start, and its level then drops as the wine is transferred to the glasses.
- Role of spreading activation for semantic priming
  - Diluted by large fan-out (vertices with many edges)
  - Compensate via splitting large graphs into overlapping smaller context related graphs, and treating context accordingly
- Verbs as causal relationships, e.g. opening and pouring
- Need to be able to model past, present and imagined situations, including plans, and what-if reasoning
  - Story telling is ubiquitous in human culture

Example Use Cases

- Computer-aided diagnosis

  A clinician will typically deal with an incomplete patient picture, and, using plausible reasoning such as abduction, generalisation, and analogy, they narrow down the many options to only several

- Computer-interpretable guidelines

  A clinical action (e.g., prescription of drug) will have an intended effect, but with a certain belief (or likelihood). Effects of drugs can also have probability distributions in time, as per pharmacology studies. Given a workflow of tasks, each affecting the patient in different ways and with different likelihoods, planning a "care path" involves searching for an optimal path towards a target state.

- Incomplete health KG

  Missing causal associations in health KG, such as between diagnosis and treatments, can be found via plausible reasoning over curated, large-scale medical knowledge. For instance, using medical hierarchies and relations, one can find the most specific body part to which both diagnosis and treatment apply.

- Literature-based discovery

  A literature-based KG, built with relations and concepts extracted from clinical literature (using NLP) is typically imperfect; by applying plausible reasoning (e.g., using word embeddings and graph traversal), the identification of missing relations (link prediction) may lead to the discovery of previously unknown causal links.

- Activity recognition

  Knowledge based approaches to human activity recognition that can cope with uncertainties arising from sensors and *idiosyncratic individual behaviours.*

- Safety argumentation

  Safety cases are essentially a structured argument supported by evidence and intended to show that a system is acceptably safe when used in a specified way. It shows how you are managing the risks as far as it is practical.

- Ethical argumentation

  Ethical review of an AI application as a structured argument that considers different lines of reasoning bearing on whether or not the use of AI conforms to accepted guidelines.

- Cybersecurity assistant

  A cybersecurity assistant can support human security specialists by monitoring 24x7 for anomalous behaviour and policy violations, and taking immediate action according to policies set by the humans, which can include alerting the on-duty human security specialists. The actions may include using software defined networking to isolate suspicious traffic and to mitigate denial of service attacks.

- Digital angel

  Cognitive agents that act as digital angels tracking our personal data across the Internet and checking whether proposed operations and uses are consistent with our personal values and preferences. This is inspired by the WEF 2022 report Advancing towards digital agency - the power of data intermediaries.

Taken from GitHub Issues

https://github.com/Imperfect-Knowledge/ik2022/issues

- Use cases and requirements
- Qualitative vs Quantitative metadata
- Causal knowledge & reasoning
- Queries and dialogues
- System 1 and 2
- Natural Language
- Neurosymbolic AI
- Systematic biases
- Graph of overlapping graphs
- Better than us
- Positives & negatives, open & closed ranges
- Variables
- Spatio-temporal reasoning
- Qualitative and fuzzy reasoning
- Social reasoning
- Relationships
- Inductive reasoning
- Cognitive architecture