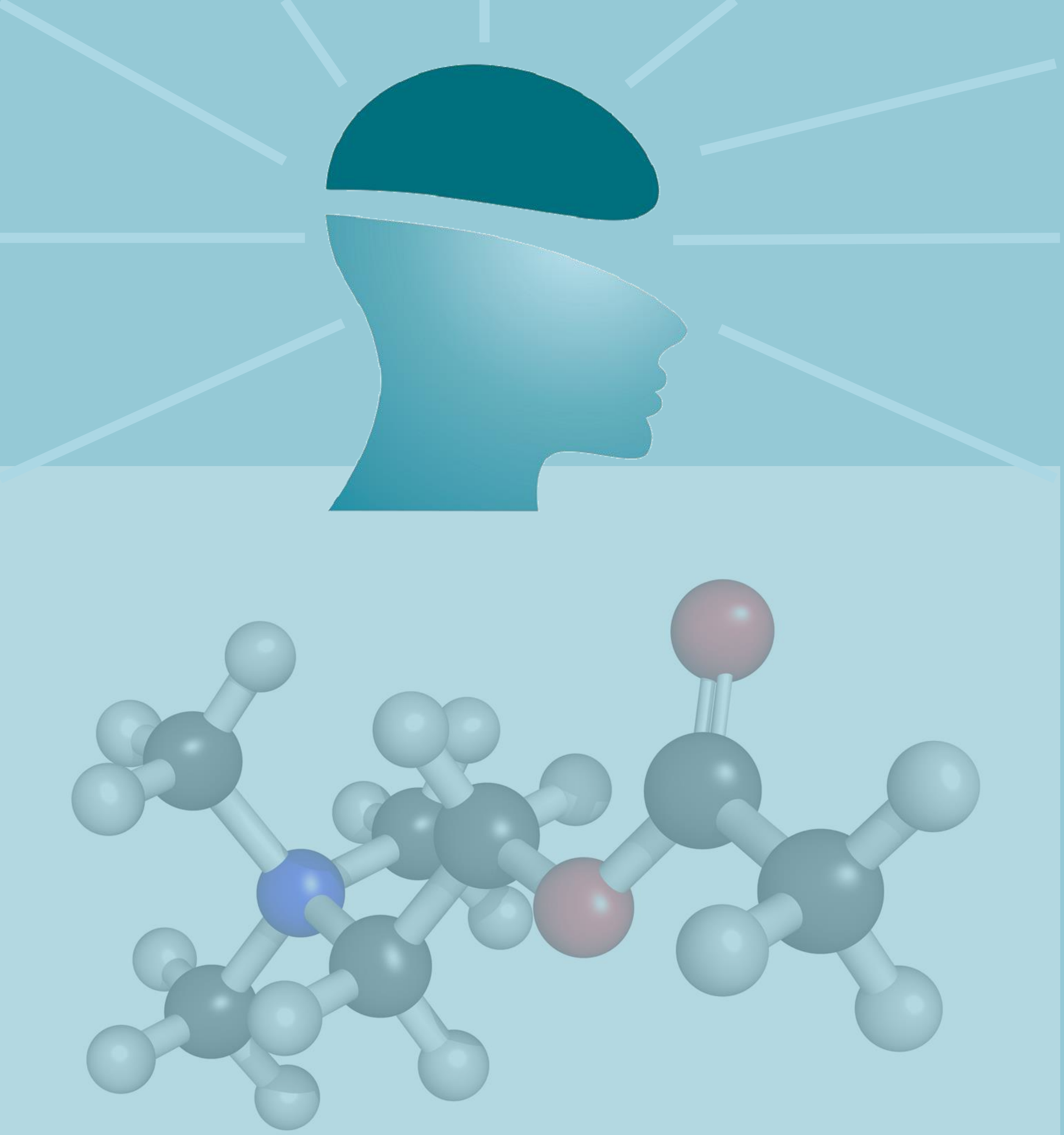
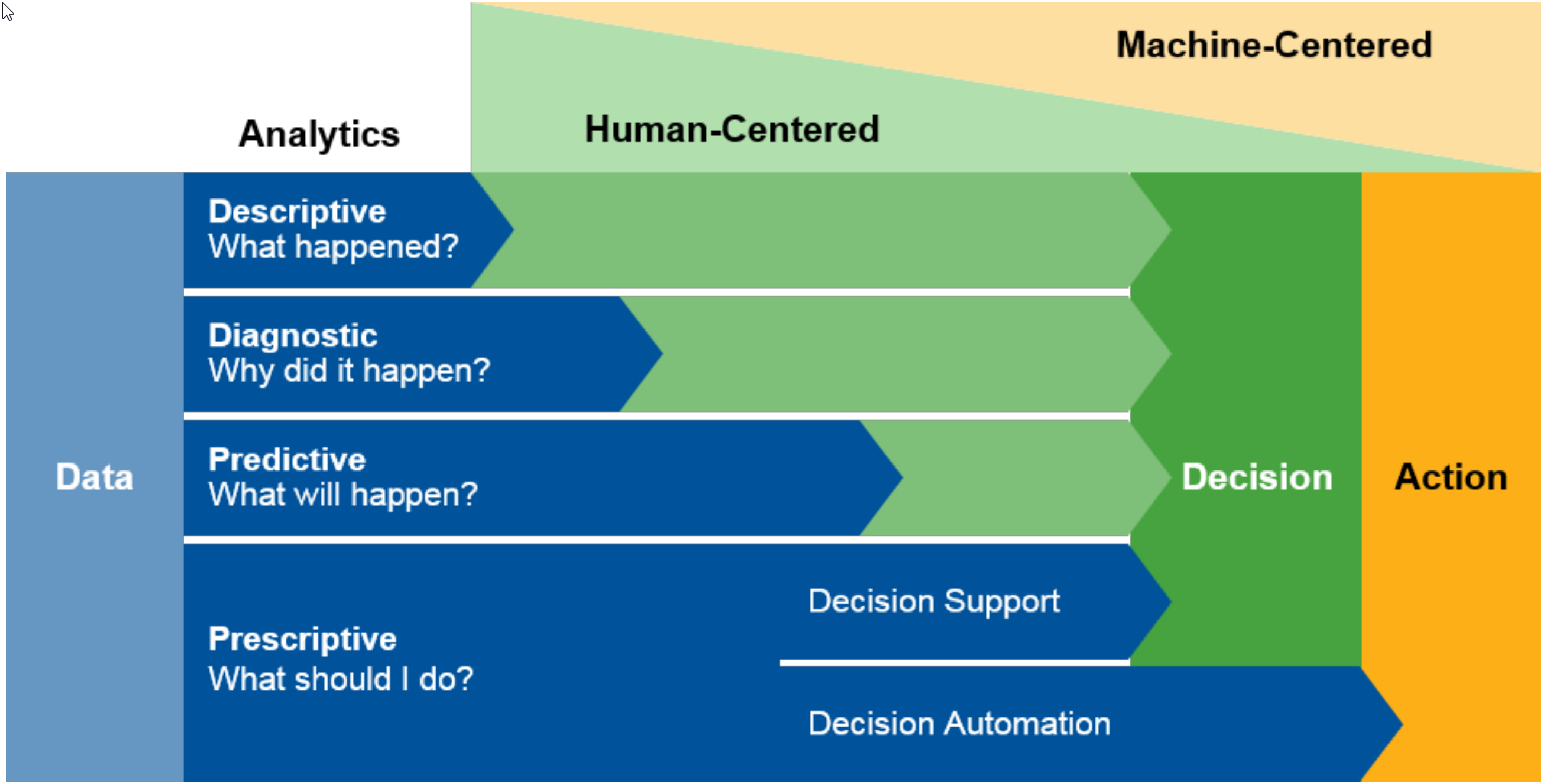


Inside the Numbers:

Architecting Decision Support with Causality to Explain Trends and Outcomes



Source: Gartner (October 2016)

Why do you want inside the numbers?

- Decision support and automation need causality
- Numbers can't show intent or "why" or "how"
- Human intent is critical to much AI and ML
- ” Being data-rich doesn't mean you are insight-rich”
  - Even quantitative intelligence can lead analysts astray
- Models for inferring intent from customer purchase or browsing activity are useful, but limited
- Much money can be saved by machine-centered qualitative analytics

Numbers aren't everything

How do you get inside the numbers?

- Semantic or Conceptual Enterprise Info Model
  - Indexes database tables and columns by concept
  - Indexes documents and web page paragraphs by concept
  - Links concepts in a concept graph
  - Attached to glossary and KPIs
  - Defines causal chains or paths from root cause to effect
- "Hypothetical Model" describes real world
  - Set of expectations with adjustable confidence values
- Natural Language Understanding = Words+Numbers

Models are needed for NLU and Causality

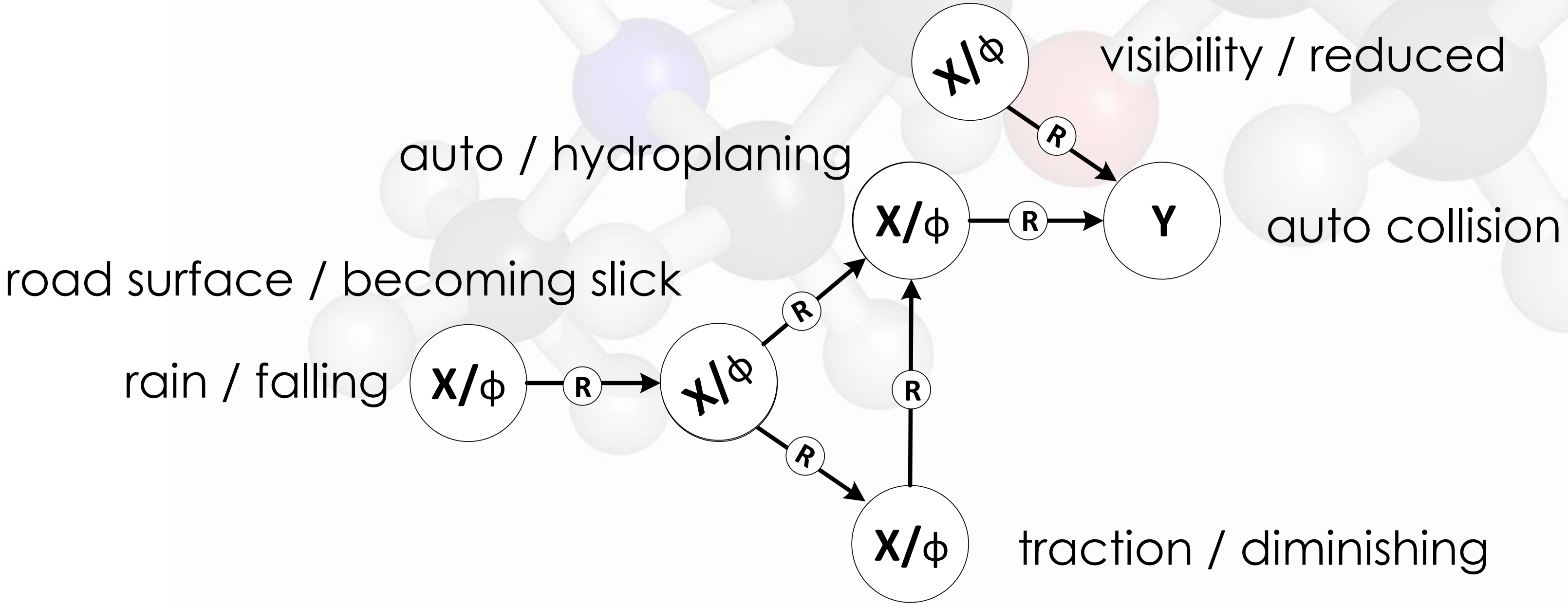
Converging Knowledge through Meaning

- Databases
  - Detail transaction level or summary – source of numbers
  - Memo fields often ignored or treated separately
- Documents
  - Word, Adobe PDF, GoogleDocs and Html pages
  - Content related to customers, business goals, products
- Ontologies
  - Graph models of taxonomies, contexts and concepts
  - Interact with indexes so you don't need to tag content

Semantic models bridge DBs and Documents

Hypothetical Models – Causal Paths

- As in Expert Systems and Decision Support Systems, the model is the basis for reasoning
- Conceptual or Semantic Graphs are ideal
- Paths traverse from root cause to outcome

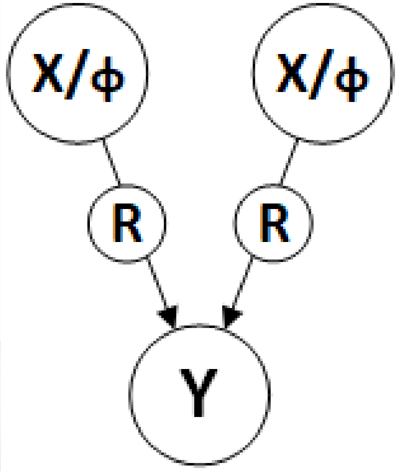


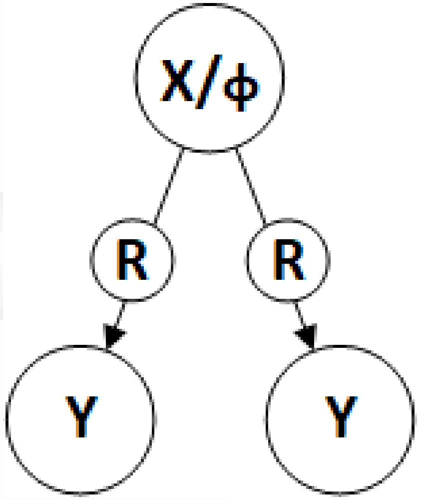
Causal Paths are Graphs

Inside the Numbers

Confounders and Colliders

- Confounders and colliders are ubiquitous
- Not knowing confounders can result in poor decisions
- Spurious correlations can impair decisions
- Not knowing colliders can result in poor predictions
- Hypothetical models enable “fill in the blanks” automation





A-priori Knowledge Needed for Reasoning

Hypothetical Models and Graph DBs

- Graph databases can optimally represent hypothetical models as semantic networks
  - Every record has Nodes and Relationships
  - Attributes Associated with both Nodes and Relations
  - Native graph traversal algorithms increase efficiency
- Concept learning can be graphically supervised
  - It is easy for humans to understand relationships between concepts, including causes and effects as linked nodes when visualized in graph database

Relational DBs and Hadoop have limits

Beyond Causal Knowledge

- While causal models are typically causal paths only causality is intrinsically connected to all knowledge
- If you know that the Sun is a star (taxonomy), you know that stars’ characteristics apply to the Sun
- If you know that a bicycle has pedals (meronomy), you can know that pedaling propels the bicycle
- If you know about time sequences, you can predict in what order events are likely to occur

Segregating knowledge hobbles reasoning

Inside the Numbers

Examples of Applicable Knowledge

Taxonomy

Time

Space

Causality

Identity

Composition

Intent

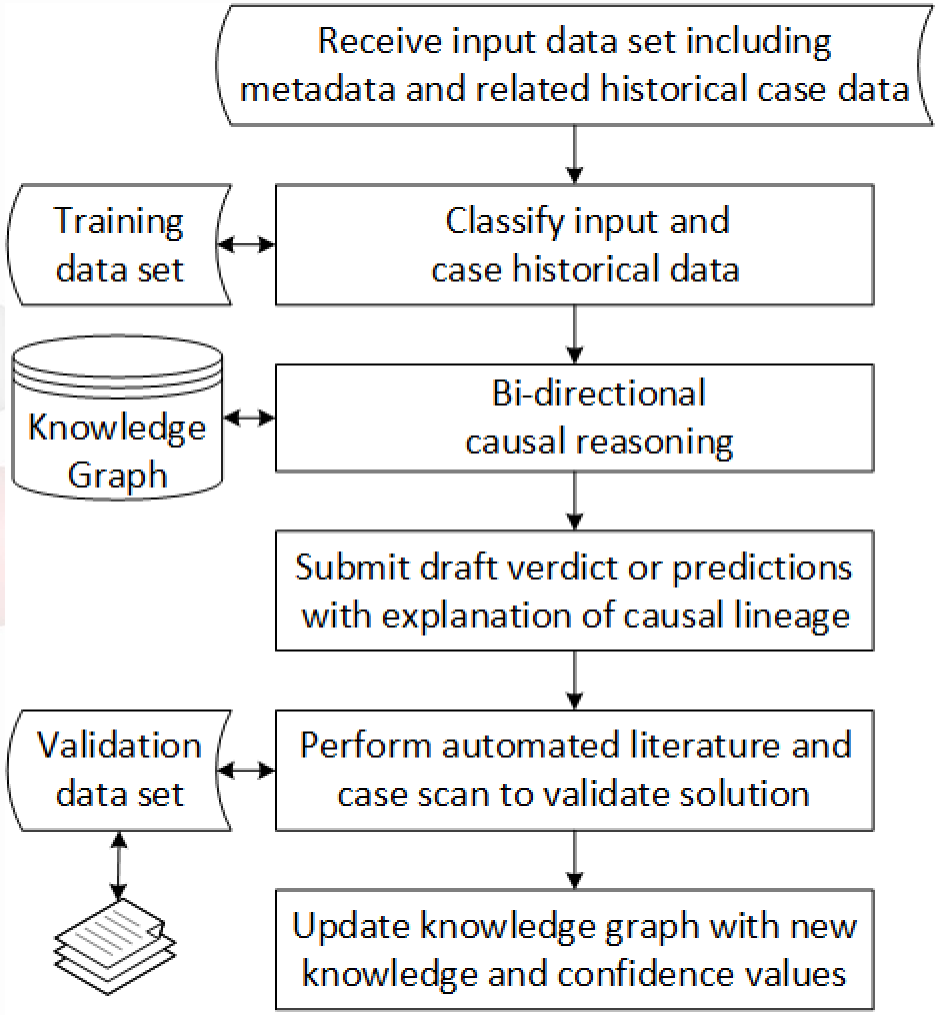
Object	Relationship	Associate	Context	Qualifier
celestial body	instance	object	universe	natural
star	instance	celestial body	cosmos	emitting light
nuclear reaction	mechanism	emitting light	star	continuous
planet	instance	celestial body	cosmos	emitting no light
orbit	motion	celestial bodies	space	constant
galaxy	group	star systems	universe	gravitationally bound
Milky Way	instance	galaxy	universe	local to humans
star system	group	celestial bodies	galaxy	gravitationally bound
Solar System	instance	star system	Milky Way	local to humans
Sun	instance	star	Solar System	central
Earth	instance	planet	Solar System	Inhabited
Earth	route	around the sun	Solar System	Earth's orbit
Earth	motion	revolving	space	around the sun
Earth	motion	rotating	space	daily
Earth revolving	causes	season change	Earth's orbit	elliptical
Earth rotating	causes	day-night cycle	solar system	24 hours
sunrise	event	day-night cycle	Earth	day's beginning
sunset	event	day-night cycle	Earth	night's beginning
sunrise	event	day-night cycle	Earth	night's ending
sunset	event	day-night cycle	Earth	day's ending

Prediction for Decision Support

Inside the Numbers

Causal Reasoning Processes

- Input and classify case and historical data
- Apply inheritance
- Search Hypothetical Models for applicable causal paths
- Infer causes or outcomes
- Draft solution with lineage
- Validate and conduct ML



Distributed heuristics/microservices can work

Inside the Numbers

Inference and Natural Language

1. User dictates whether to infer outcomes or causes
2. System is smart enough to know the difference
   - Tokenize and semantically classify data in the input
   - Find cues that differentiate causes from outcomes
   - Use contextual information to fill in blanks
   - Natural Language Understanding needed to automate
   - Without NLU, all the data must be fully tagged
   - System guesses can be machine or human validated
   - Validation results can be fed to learning algorithms

More learning → Less human input required

Inside the Numbers

Working with Incomplete Case Data

- Hypothetical causality models become more complete over time (e.g. weather forecasting)
- Taxonomy, meronomy, time, space models needed
- Interconnections between phenomena are ubiquitous
- The more models grow, the more they merge
- A single universal model is the inevitable end
- A single universal model is needed for robust NLU
- Chicken and Egg problem

Start with a seed of knowledge and accumulate

Architectural Implications

- NLU Modeling Capability or Vendor
- Minimize Core System Customizations
- Internal AI resource is a must-have for continuity

NLU and Causal Reasoning use Same Model

- Break down information silos between functional areas
- Build graph models spanning multiple functional areas

Architect Intelligent Periphery Around Core

- Select ERP, CRM, MRP, PLM, LIM with good APIs
- Define reusable BI algorithms and heuristics around core

Architect out of the box

- Commodity AI (Watson, Einstein) alone can't deliver as much value
- Ensure sufficient organic AI dev/curation resources

AI won't succeed without business commitment

Conclusion

To reduce the level of human analysis needed to deliver qualitative, actionable intelligence, architect:

- Model-based natural language understanding and causal reasoning in an intelligent periphery
- Graph-structured model with causal paths plus time, space, taxonomy, meronomy and other knowledge
- NLU, causality and learning can use the same model
- Distributed heuristics for each type of knowledge and special accommodations for colliders and such

Bibliography

- Richard O. Duda, Peter E. Hart and Nils J. Nilsson, "Subjective Bayesian methods for Rule-Based Inference Systems", vol. 2, 1976, pp. 192-199, 212-229.
- "Qualitative Reasoning About Physical Systems: An Introduction", D.G. Bobrow, Artificial Intelligence 24, pp. 1-6, 1984.
- Tom Bylander, "CRSL: A Language for Classificatory Problem Solving and Uncertainty Handling", The AI Magazine 8/1986, pp. 66-77.
- Carlson et al., "Frame-Based Modeling Techniques for Analyzing A Strategic Planning Domain", Proceedings of the Twenty-First Annual Hawaii International Conference on System Sciences vol. III Decision Support and Knowledge Based Systems, 5-8 Jan. 1988, pp. 131-135.
- Judea Pearl, “The Book of Why: The New Science of Cause and Effect”, Basic Books, 2018.