[Journal of the Association for Information Systems](https://aisel.aisnet.org/jais)

Volume 25
Issue 1 *Special Issue: The Future Impact of AI on Academic Journals and the Editorial Process (pp. 37-181)*

Article 3

2024

Extending the Foresight of Phillip Ein-Dor: Causal Knowledge Analytics

Richard T. Watson
Digital Frontier Partners / University of Georgia, richard.watson@digitalfrontierpartners.com

Yuanyuan (April) Song
University of Georgia, yuanyuan.song@uga.edu

Xia Zhao
University of Georgia, xia.zhao@uga.edu

Jane Webster
Queen's University, jane.webster@queensu.ca

Recommended Citation

Watson, Richard T.; Song, Yuanyuan (April); Zhao, Xia; and Webster, Jane (2024) "Extending the Foresight of Phillip Ein-Dor: Causal Knowledge Analytics," *Journal of the Association for Information Systems*, 25(1), 145-157.
DOI: 10.17705/1jais.00871
Available at: https://aisel.aisnet.org/jais/vol25/iss1/3

Journal of the Association for Information Systems (2024) **25**(1), 145-157
doi: 10.17705/1jais.00871

EDITORIAL

ISSN 1536-9323

Extending the Foresight of Phillip Ein-Dor: Causal Knowledge Analytics

Richard T. Watson,[1] Yuanyuan (April) Song,[2] Xia Zhao,[3] Jane Webster[4]
[1]Digital Frontier Partners, Australia / University of Georgia, USA, richard.watson@digitalfrontierpartners.com
[2]University of Georgia, USA, yuanyuan.song@uga.edu
[3]University of Georgia, USA, xia.zhao@uga.edu
[4]Queen's University, Canada, jane.webster@queensu.ca

Abstract

Phillip Ein-Dor advocated that electronic journals be more than a PDF of the established text model. He envisioned a transformation of scholarship. The need for such a transition has only grown since the first issue of JAIS in 2000 because the continuing growth and fragmentation of knowledge limits the generation of new knowledge. We propose drawing on analytics and AI to accelerate and transform scholarship, providing an appropriate tribute to a visionary IS leader.

Keywords: Knowledge, Graph Database, Causal Knowledge Analytics, Network Analysis, Natural Language Processing, Artificial Intelligence, Large Language Models, Machine Learning, Knowledge Engineering

Dov Te'eni was the accepting senior editor. This paper was submitted on April 29, 2023 and underwent two revisions. It is part of the Special Issue on The Future Impact of AI on Academic Journals and the Editorial Process.

1 Electronic Journals and Knowledge Engineering

As the foundation editor of the *Journal of the Association of Information Systems* (JAIS), Phillip Ein-Dor[1] was a leading proponent of electronic journals. His vision of scholarly publication was to "transform the evaluation of research publications from one with few participants to one in which everyone interested can participate and the best papers will survive" (Peffers et al., 2003, p. 501). We align with his goal to change the assessment of research by enhancing authors' analytical capabilities to broaden their comprehension of a research domain. We suggest a path aligned with his vision by complementing current literature reviews with computational analytics.

Phil recognized that technological means are required to manage knowledge. He advocated knowledge engineering (Ein-Dor, 2011, p. 1490). He argued that electronic journals should extend print beyond a PDF facsimile to incorporate links to other knowledge sources. We can see examples of this in the early editions of JAIS. Phil recognized that to analyze knowledge, one must first select a form of representation. We agree with his contention and, in so doing, add another category, *causal knowledge*, to his eight dimensions of knowledge (Ein-Dor, 2011). We propose that a digital representation of an article's causal model or propositions, when present, provides a

[1] Phillip Ein-Dor was born in Australia, and in keeping with the informality of Australia, we refer to him as Phil in this publication. As a modest person known to everyone as Phil, we are sure this casualness is not misplaced nor disrespectful of an admirable scholar and AIS leader.

foundation for a new branch of knowledge engineering. This extension, *causal knowledge analytics*, applies methods such as graph analytics and artificial intelligence to digitized causal models. Thus, we pay homage to Phil's pioneering vision for electronic journals and his prescient identification of the need for knowledge engineering, which is "the branch of engineering that analyses knowledge about some subject, and transforms it into a computable form for some purpose" (Sowa, 1999). We implement knowledge engineering by transforming visual causal models or textual propositions into graphs and illustrate how a knowledge graph dataset affords five classes of computable forms. Our approach achieves the following goals: it (1) complements a traditional text-driven synthesis of knowledge with metrics and visualizations, (2) raises scholarly productivity because a knowledge graph dataset is shareable, (3) supports reproducible analytics because a set of queries and computational methods can be rerun for a different domain or the same domain some years forth, and (4) lays the foundation of building causal analytics as a research methodology.

2 Knowledge Accumulation

Knowledge accumulation was slow and splintered until it began to be institutionalized in the middle of the 17th century through peer reviewed journals (Bornmann & Mutz, 2015). For the first 100 years, scientific publications grew at about 1% per year. In the early 21st century, growth rates reached 34% per year (Michels & Schmoch, 2012). However, the success of the knowledge creation system has ironically been threatened by its accelerating growth (Park et al., 2023). As a result, the average age at which leading scientists and inventors produced their most valuable work rose by six years during the 20th century. New scholars take longer to reach the frontier of their field, and knowledge creators' working lives are shorter (Jones, 2010). Scholars need new digital tools to accelerate reaching their discipline's front lines (Matthews, 2021).

The recognition of the need to support knowledge searching (Garfield, 1955) has resulted in citation indices for the natural and social sciences. Google Scholar is a digital enhancement of these indices. However, these tools identify potential knowledge sources but do not encode knowledge in a digital format (Larsen et al., 2020). Tools that support IS literature searching (e.g., Boell & Wang, 2019) usually rely on "points of access," such as title words, key terms, and classifiers. Such access paths rarely encode the relevance of a publication to current and future problems or theories (Swanson, 1986). The lack of standard digital encoding of knowledge inhibits the growth of many fields. New digital tools are needed to accelerate scholarship by structuring existing knowledge to facilitate search, analytics, and the computational capabilities envisioned by Phil (Ein-Dor, 2011, p. 1490).

We propose a path for the emergence of *causal knowledge analytics* (CKA), defined as *the development and application of methods for processing graphical representations of causal models to advance theoretical research by scholarly communities.* We address four questions: namely (1) what knowledge to codify, (2) how to represent this knowledge, (3) how to analyze codified knowledge, and (4) how to interpret these analyses. Before considering these questions, we examine prior work on the meaning and management of knowledge.

3 Knowledge

Most human knowledge is grounded in causal explanations and the mental models we create to make sense of reality (Pearl & Mackenzie, 2018, p. ch. 1). Such knowledge can be represented in graphical form, which, if standardized, can facilitate the connection of related concepts across publications. High-quality codified knowledge is a prerequisite for computational analytics. Unless independent causal chunks are logically connected, their relationships are essentially invisible. For example, unconnected public fragments might report A → B and B → C, but unless they are linked through B, the knowledge that A → C may remain undiscovered (Swanson, 1986).

4 Knowledge Codification

While causal models do not capture all knowledge, they make much of it amenable to communication and in-depth investigation. A causal model represents some aspect of the world's causal structure by specifying how (and if) concepts or variables are related (Pearl, 2009). A causal model expresses what is known or conjectured through boxes and arrows. When digitized, causal models can be analyzed with computational methods and natural language processing (NLP).

Causal models, standardized as graphs with nodes (concepts) and edges (relationships), provide scholars with a shared lexicon. They are a means by which a community of inquiry can link concepts and relationships for constructing meaning. As a unifying abstraction, graphs collectively create an opportunity to explore and analyze scholars' interpretations of the world. Metadata standards, such as one for causal graphs, increase the value of technology for searching and analyzing (Musen, 2022).

Formal representations are more amenable to analysis because codifying knowledge reduces ambiguity, but the process precludes the nuances embedded in text. Despite this important caveat, our stance is that those fields that apply computer-based methods to codified knowledge will lower the threats of fragmentation and research deceleration (Kohlhase, 2010). Moreover, with sufficient coded causal models, we will have a "tagged" dataset appropriate for applying machine learning (Negro, 2021) using graph transformers (Koncel-Kedziorski et al., 2019). A comprehensive graph database enables an exploration of the literature that moves beyond the intrinsic shortcomings of current large language models (LLMs), arguably the most advanced form of general AI. LLMs cannot support "precise mathematical calculation, multi-step logic reasoning, perception about the spatial and topological factors, and handling the temporal progression" (Zhang, 2023).

5 Coding Causal Knowledge

Knowledge is built upon a foundation of inherited wisdom and supported by a framework of established principles. We propose capturing this in a data model for a publication (Figure 1), recording a causal model's theoretical foundations, concepts and their definitions and relationships, authors, and citation information. Concepts frame a causal model by using relationships to form a graph's topology. A concept, however, unaccompanied by a definition, leaves each reader to make assumptions about the author's meaning. Literature reviews often report multiple definitions of a concept, such as Vial's (2019) 23 different definitions of digital transformation. Thus, the set of concept definitions on which a causal model is based must be captured.

Knowledge evolves. It descends from prior research and is selectively modified over time to describe contemporary phenomena. Coding recent casual models does not ignore the past because prior knowledge is passed down through theories, concepts, and relationships. Scholars are expected to acknowledge the theories that shape their thinking (Locke & Golden-Biddle, 1997), and these theories should be recorded. This audit trail is critical information for new scholars joining the current discourse stream and for those embarking on extending current knowledge.

5.1 Digitizing Causal Models

We have designed a data model (Figure 1) to demonstrate the possibilities of causal knowledge analytics. It records that a publication has authors, depicts causal models, is built on theories, and contains relations between elements and elements' definitions.[2] A causal relation contains elements and can be derived from theories. A theory can refer to other theories, inform elements, and propose relations. We visualize a publication as shown in Figure 2.

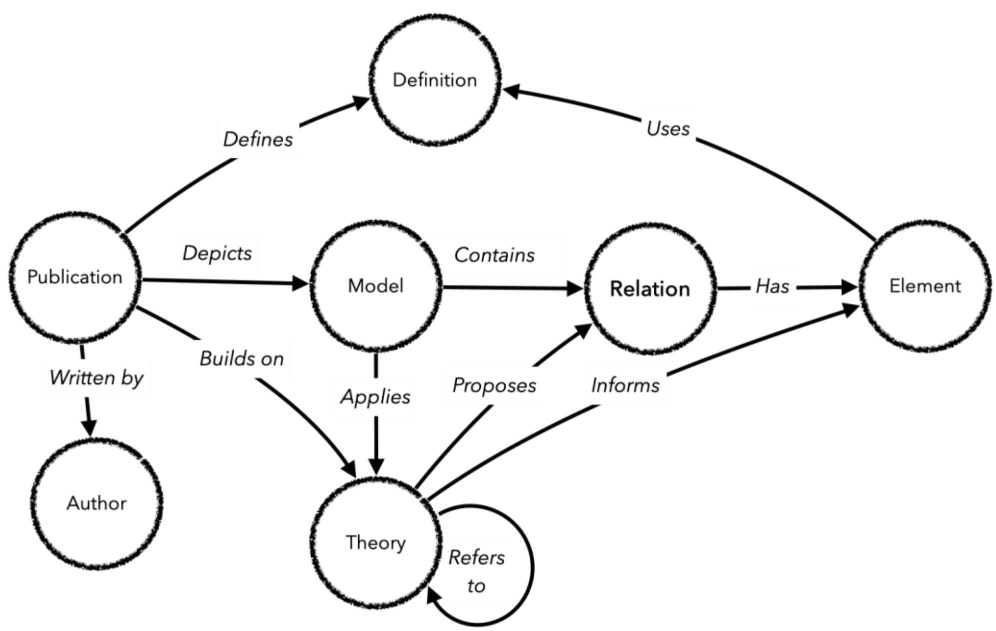


Figure 1. A data Model for Causal Knowledge (Song et al., in press)

[2] Element is a general term for constructs, concepts, and events.

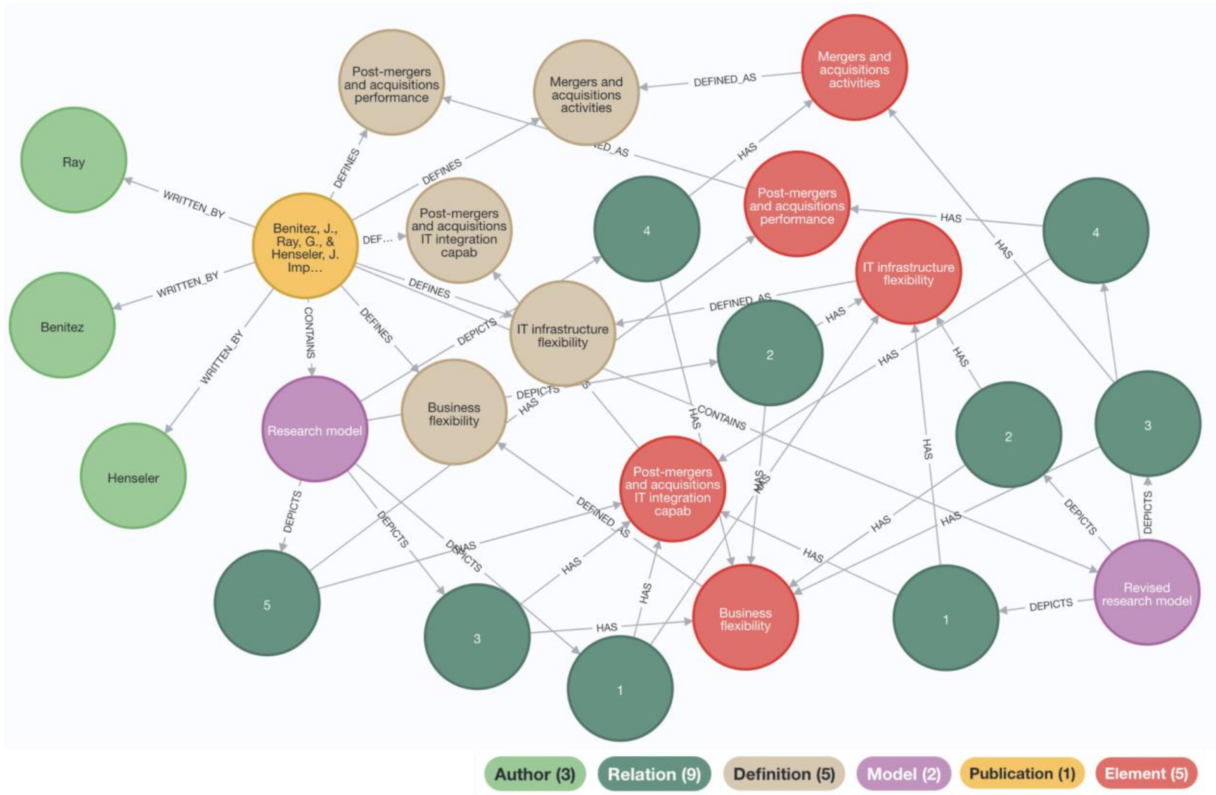


Figure 2. A Graphical Representation of an *MIS Quarterly* Curation Publication (Benitez et al., 2018)

6 Causal Knowledge Analytics

Millions of knowledge miners and creators need tools to increase their efficiency. While there are useful frameworks for literature reviewing (Templier & Paré, 2015; Wagner et al., 2021), it is time for scholarship to emulate the many organizations that have adopted digitization to facilitate new forms of analytics. Codification can leverage diverse computational methods to explore a knowledge network and expand scholars' capabilities for processing literature. It also creates a foundation, by "tagging" knowledge for using next-generation LLMs to create knowledge about knowledge.

6.1 Opportunities for Applying Causal Knowledge Analytics

We propose five levels of analysis (Table 1), which gradually extend the breadth of the view of analysis. We identify some methods for addressing these main types of analyses, and we expect that future research will surface more. We now illustrate possible analyses and comment on their potential value.[3] The framework identifies levels of analyses for a complementary approach to literature reviewing based on network metrics and LLM textual processing. Thus, the identified methods include computational techniques (e.g., social network analysis) and textual analytics (e.g., BERT).

Element level: At the element level, we can retrieve information and measure the semantic similarity of elements using NLP methods. It is quite common to have multiple definitions for the same concept. Understanding how these definitions vary is fundamental to understanding a domain. Using BERT (Devlin et al., 2018), we can compute the similarity of a concept's definitions (Table 2). The most similar definitions are 1 and 3 and 1 and 5 (bold), while the least similar are 1 and 7 and 5 and 7 (italics) (Table 3).

Relationship level: At the element-relationship level, a graph database can connect concepts and relationships reported independently in publications. Queries can reveal antecedents or consequents of an element and their frequencies represented as the number of edges between two concepts (Figure 3). Causal paths can be reported starting with an element, ending with an element, or between two elements (Figure 4). Querying a graph database is less burdensome than reading and manually synthesizing causally related concepts.

[3] The illustration of the five-level analysis is based on 238 papers with an explicit causal model in 13 *MIS Quarterly* curations (https://misq.umn.edu/research-curations; https://t-rex-graph.org/database).

Table 1. Five Possible Levels of Analysis of a Knowledge Graph

Level of analysis	Tasks	Examples	Possible methods
Element (concept or event)	Information retrieval	Retrieval of information about an element, such as its definitions	Graph query language (GQL)
	Element semantic similarity	The similarity between two elements based on their definitions	NLP, such as bidirectional encoder representations from transformers (BERT) (Devlin et al., 2018); LLMs
	Jingle and jangle fallacies*	The calculation of similarities to identify potential jingle (Thorndike, 1913) and jangle (Kelley, 1927) fallacies.	NLP-based similarity matrix; BERT (Devlin et al., 2018); LLMs; machine learning methods (Barlaug & Gulla, 2021); Network similarity measures
Relationship	Information retrieval	Retrieval of related elements, such as their antecedents	GQL; graph theory methods, such as shortest path analysis
	Network importance	The importance of an element in a knowledge network	Social network analysis metrics, such as centrality
	Mediator and moderator identification	Identifying concepts used as mediators	GQL
	Literature gap	Identification of literature gaps	Social network analysis; link prediction methods (e.g., Zhang & Chen, 2018) and network measures (e.g., Katz, 1953); graph neural networks (Zhou et al., 2020)
Model	Jungle conundrum	Identification of similar causal models	Graph isomorphism analysis (Song et al., 2021)
	Endogeneity issues	Identification of potential endogeneity issues	DAG analysis (Textor et al., 2011)
	Concept co-occurrence analysis	Identifying concepts that appear together in a model	Association rule analysis
Theory	Identification of theories	Retrieve theories and related topics	GQL
	Theory impact	Analysis of the impact of a theory	GQL to report the frequency of a theory's application; network metrics to analyze theory use.
Theme	Knowledge fragmentation	Evaluation of the cumulative nature of a knowledge network	Social network density
	Comparative analysis	Tracking the development of a theme	Cluster analysis (Shmueli et al., 2017)
	Model integration or simplification	Condense and simplify causal models related to a particular topic	Graph summarization (Liu et al., 2018)
	Causal consensus analysis	Computing the consensus of relationships	Information entropy (Shannon, 1948)
	Theme identification	Identification of research themes	K-means, NLP, or LLMs for clustering; community detection (Chen et al., 2017; Fortunato, 2010)
	Element-topic correlation analysis	The association of elements and research topics	Correspondence analysis
	Knowledge development	Analyzing the evolution of a concept's network measures	Network analysis, such as structural hole theory

Note: *The jingle fallacy occurs when two concepts have identical or similar labels but reference different real-world phenomena. The jangle fallacy occurs when two concepts reference identical or similar phenomena but are labeled differently.

Table 2. Definitions of Trust

	Definition
1	"We define trust as the subjective assessment of one party that another party will perform a particular transaction according to his or her confident expectations, in an environment characterized by uncertainty" (Ba & Pavlou, 2002, p. 245)
2	"Trust is argued to be rooted in perceptions of teammates' ability, benevolence, and integrity (Jarvenpaa et al. 1998). Ability refers to the aptitude and skills that enable an individual to be perceived as competent by teammates (Jarvenpaa et al. 1998; Mayer et al. 1995). Benevolence refers to the extent to which an individual is believed to be willing to help teammates beyond personal motives or individual gain. 1995). Integrity refers to the extent to which an individual is believed to adhere to a set of principles thought to make her dependable and reliable" (Piccoli & Ives, 2003, p. 366).
3	"Trust is defined as the buyer's intentions to accept vulnerability based on her beliefs that the transaction will meet her confident expectations" (Pavlou et al., 2007, p. 107)
4	"The user beliefs in the recommendation agents' competence, benevolence, and integrity. The beliefs that 1) the recommendation agent has the ability, skills, and expertise to perform effectively 2) the recommendation agent cares about the user and acts in the user's interest 3) the recommendation agent adheres to a set of principles (e.g., honesty and promise keeping) that the user finds acceptable" (Xiao & Benbasat, 2007, p. 144).
5	"Trust reflects one party's belief that its requirements will be fulfilled through future actions undertaken by the other party" (Goo et al., 2009, p. 126).
6	"Trust is conceptualized as a single variable and refers to general confidence in the website" (Cyr et al., 2009, p. 545).
7	"The extent to which a buyer perceives in a seller's ability (i.e., skills, competencies, and characteristics in seller his/her products online), integrity (adhering to a set of principles that the buyer finds acceptable), and benevolence (i.e., doing good toward the buyer)" (Ou et al., 2014, p. 217).

Table 3. Concept Similarity for Trust Definitions

Definition	1	2	3	4	5	6	7
1	—	—	—	—	—	—	—
2	0.60	—	—	—	—	—	—
3	**0.80**	0.63	—	—	—	—	—
4	0.57	0.76	0.63	—	—	—	—
5	0.81	0.61	**0.77**	0.57	—	—	—
6	0.59	0.50	0.57	0.49	0.46	—	—
7	*0.38*	0.57	0.52	0.76	*0.38*	0.45	—
Means	0.62	0.61	0.66	0.63	0.60	0.51	0.58

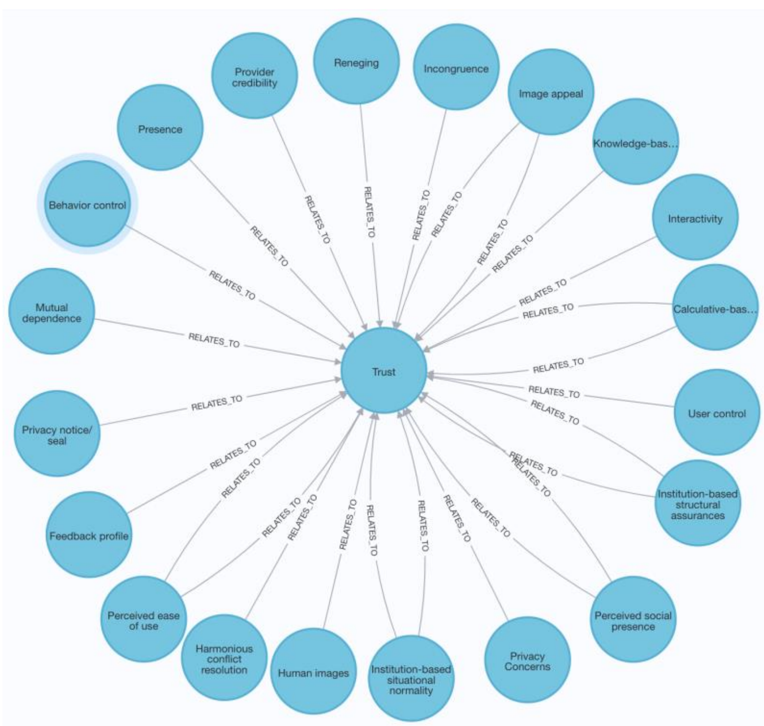


Figure 3. Antecedents of Trust

Figure 4. Illustration of a Causal Chain

Model level: A knowledge network offers two possibilities for scientific advancement. First, new models may emerge from analyzing cause-effect relationships reported in different publications. Because models are fragmentally embedded in various publications, retrieving pieces and integrating them to construct a current state of knowledge model is challenging. Graph isomorphism analysis can report similar models (Cordella et al., 2004) (Figure 5). Second, reported models will typically vary in their similarity, and assessing their degree of difference can help researchers study the evolution of a topic.

Theory level: We can identify theories that inform an element in one or more publications, spotlighting various theoretical perspectives of a concept. A starting point is to investigate the application of theories by querying the topics associated with theories (Table 4).

Theme level: A theme-level analysis provides a holistic view. Measuring concept density can evaluate and compare knowledge fragmentation in different knowledge networks. Social network analysis can identify structural holes and weak ties. A trend analysis of structural holes based on effective size (Burt, 1992) can show how a field develops. For illustration, we compare two knowledge networks in Table 5 (1983-1995 and 1995-2020) and report the effective size changes, which indicate that concepts such as *perceived usefulness* and *perceived ease of use* have gained prominence since 1995.

6.2 The Complementary Role of LLMs

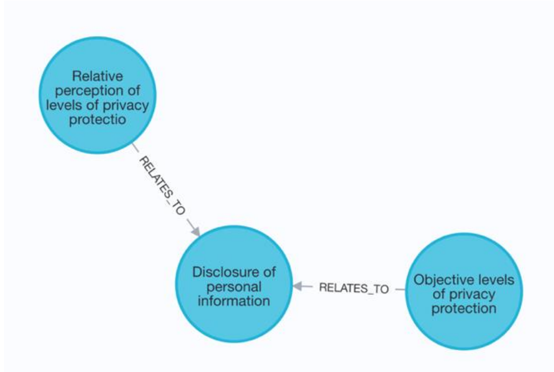
The key differences between LLMs and CKA (Table 6) reveal that they are complementary approaches to improving literature reviewing, reporting the nature of existing knowledge, and exploring causal connections. We consider these differences using an input-processing-output (IPO) approach.

6.2.1 Input

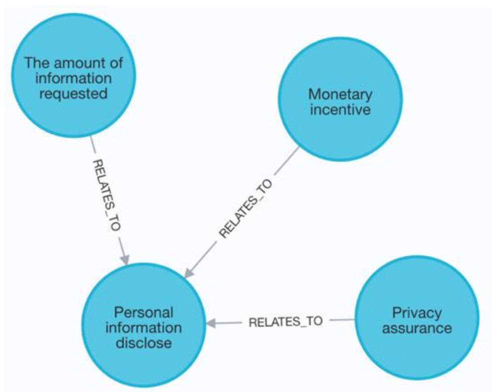
LLMs process unstructured digitized text files and PDFs. Text is inherently ambiguous at all levels of linguistic analysis (Piantadosi et al., 2012). The frequency of syntactic and semantic ambiguity is a significant challenge to NLP. Furthermore, current LLMs do not model ambiguity (Liu et al., 2023). The essential structures and characteristics of language did not seemingly evolve for precise communication (Chomsky, 2002, p. 107). To apply an old IS adage, *ambiguity in—ambiguity out*.

In contrast, standardized coding and measurement are a means for reducing ambiguity. The metric system replaced thousands of local measurement systems, which reduced ambiguity in trading.[4] Databases are designed to reduce ambiguity by defining entities, relationships, and attributes. This approach makes databases powerful platforms for operational management and analytics. Databases do not eliminate ambiguity. Fields containing text strings, such as a concept's definition, can be ambiguous because they are based on everyday language.

[4] Prior to the adoption of the metric system, France had over 250,000 local measurement variants (Zupko, 1990).



Adjerid et al. (2018)



Hui et al. (2007)

Conceptual isomorphism score = 0.696

Figure 5. Similar Models for Trust

Table 4. An Example of Some Common Theories and Related Topics (Curations)

Theory	Curations	Number of curations
Agency theory	IS Control and Governance; IS Sourcing; Information Privacy; IT Project Management; Trust	5
Theory of planned behavior	IS Control and Governance; IT Workforce; IT Project Management; Securing Digital Assets; Trust	5
Adaptive structuration theory	IT-supported Collaboration; Health Information Technology; IS Use; Trust	4
Institutional theory	Health Information Technology; IS Sourcing; IS Use; Securing Digital Assets	4
Prospect theory	IS Control and Governance; IT Workforce; Information Privacy; IT Project Management	4

Table 5. Concepts with Highest Effective Size Changes

Concept	Effective size change
Perceived usefulness	17.9
Perceived ease of use	14.9
Performance	11.5
Behavioral intention	9.1
Job satisfaction	8.4
IT use	7.5
Social presence	7.0
Attitude	5.8

Table 6. Key Differences between LLMs and CKA

Characteristics	LLMs	CKA
Input	Unstructured text	Standardized data
Computation	No support for computing relationship metrics	Wide variety of methods of relationship metrics
Information processing methods	NLP and LLM	Graph query language, social network analysis, information theory, association analysis, correspondence analysis, NLP, and extensible to include new methods, such as LLM.
Output	Article summarization	Concept-centric synthesis

6.2.2 Processing

The nature of an input determines how it might be processed. Clearly, the range of processing possibilities is higher with structured data because some fields are suitable for direct computation or counts. Furthermore, text fields in a database can be processed using NLP and LLM methods. A database is designed to support analytics, and thus many information processing options are available. Machine learning-based graph mining methods are facilitated by graph-represented causal knowledge (Table 1). Link prediction methods can suggest undiscovered causal relationships (Zhang & Chen, 2018). Community detection methods can cluster concepts and identify research themes (Chen et al., 2017). The fundamental difference is that an LLM "does not compute" and was not designed for mathematical manipulations. CKA is based on well-established metrics and analysis methods, while the performance of LLMs is highly dependent on the input text.[5] Although augmenting the volume of input data for LLMs could improve their performance, this approach is a computationally expensive strategy.[6]

6.2.3 Output

Because CKA is based on a graph of concepts and their relationships, it intrinsically supports a synthesizing concept-centric literature review (Webster & Watson, 2002), which is generally accepted as more appropriate than an author-oriented summarization. LLMs can summarize a collection of articles, but the integration of knowledge across many sources may include hallucination or confabulation because LLMs lack reasoning. Synthesis is a higher-level human task than summarization because it involves understanding a phenomenon from diverse perspectives. CKA can support the creation of a diverse perspective, but humans ultimately have to make sense of a set of CKA analytics, LLM outputs, and relevant original texts, including those not having a causal model, to document a comprehensive understanding of a topic. These approaches are complementary methods for discovering what is known in order to investigate the unknown. Scientific advances are built upon cumulative contributions and complementary methods.

6.3 Discussion and Conclusions

Analyzing what is known is an essential phase of scholarship. It helps new scholars become immersed in a knowledge stream and seasoned scholars to direct its flow. We can improve the discovery process by digitizing core knowledge and by providing enhanced search and analytic tools. We have introduced a five-level framework and demonstrated its use for each level. This starting point has the potential to sprout new analytical methods. Progress in AI, graph analytics, and other algorithmic methods will feed future improvements.

Although our proposed method presents many advantages for knowledge engineering, there are two important limitations. First, we have focused on coding explicit causal models (and propositions potentially expressible as causal models) to illustrate the potential of CKA. Importantly, more than 70% of JAIS articles in the last nine years fit into these two categories. For the remaining articles, our data model supports partial coding (e.g., theories applied) to comprehensively record a journal's publications. Digitization need not exclude articles without explicit causal models or propositions. Authors can code relationships based on the descriptions in their articles. As we argued earlier, publications are cumulative and inherit prior knowledge, regardless of their epistemological bent. The proposed coding method can support recording all the non-graph data about a publication, including key information such as theories and concept definitions from qualitative research. Thus, this first limitation of the current implementation is addressable.

[5] https://www.economist.com/business/2023/08/13/ai-is-setting-off-a-great-scramble-for-data

[6] https://spectrum.ieee.org/deep-learning-computational-cost

The second limitation is the cost of historical completeness—that is, coding all papers prior to the introduction of digitization. Because a new theory is crafted from earlier work (Rivard, 2020), coding recent casual models does not ignore the past. As explained, prior knowledge is passed down through theories, concepts, and relationships. Nevertheless, there may be an incomplete historical record. A partial solution would be to encourage those coding causal models to support literature reviews to submit their work to an AIS community graph database. This would be slow and incomplete but appears the only alternative until LLMs master knowledge recognition.

The current literature reviewing method of reading a set of possibly related texts is a relic in the digital age. Coding causal knowledge adds a modicum of time to the publication process. It should improve the quality of journal submissions, subject causal models to rigorous review, and ensure standardized and high-quality input for a knowledge store. Asking authors to code their causal models is like a checklist, which is a widely applied tool (Borchard et al., 2012) for ensuring process completeness—for example, when submitting a manuscript for review.

As the founding editor of the *Journal of the Association of Information Systems*, Phillip Ein-Dor initiated a digital publication journey that has evolved slowly since its first issue. It is time for another significant advance that fits the pattern of Phil’s vision. It is time for JAIS to again lead the IS field by pioneering the digitization of causal models to enable the emergence of causal knowledge analytics.

Acknowledgments

This research is supported by the Alfred P. Sloan Foundation.

References

Adjerid, I., Peer, E., & Acquisti, A. (2018). beyond the privacy paradox: Objective versus relative risk in privacy decision making. *MIS Quarterly*, *42*(2), 465-488.

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, *26*(3), 243-268.

Barlaug, N., & Gulla, J. A. (2021). Neural networks for entity matching: A survey. *ACM Transactions on Knowledge Discovery from Data*, *15*(3), 1-37.

Benitez, J., Ray, G., & Henseler, J. (2018). Impact of information technology infrastructure flexibility on mergers and acquisitions. *MIS Quarterly*, *42*(1), 25-43.

Boell, S., & Wang, B. (2019). www.litbaskets.io, an IT artifact supporting exploratory literature searches for information systems research. *Proceedings of the 2019 Australasian Conference on Information Systems*.

Borchard, A., Schwappach, D. L., Barbir, A., & Bezzola, P. (2012). A systematic review of the effectiveness, compliance, and critical factors for implementation of safety checklists in surgery. *Annals of Surgery*, *256*(6), 925-933.

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215-2222.

Burt, R. S. (1992). *Structural holes: The social structure of competition*. Harvard University Press.

Chen, Z., Li, X., & Bruna, J. (2017). *Supervised community detection with line graph neural networks.* arXiv. https://arxiv.org/abs/1705.08415

Chomsky, N. (2002). An interview on minimalism. In A. Belletti & L. Rizzi (Eds.), *On nature and language* (pp. 92-161). Cambridge University Press.

Cordella, L. P., Foggia, P., Sansone, C., & Vento, M. (2004). A (sub)graph isomorphism algorithm for matching large graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(10), 1367-1372.

Cyr, D., Head, M., Larios, H., & Pan, B. (2009). Exploring human images in website design: A multi-method approach. *MIS Quarterly*, *33*(3), 539-566.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv. https://arxiv.org/abs/1810.04805

Ein-Dor, P. (2011). Taxonomies of knowledge. In *Encyclopedia of Knowledge Management, Second Edition* (pp. 1490-1499). IGI Global.

Fortunato, S. (2010). Community detection in graphs. *Physics reports*, *486*(3-5), 75-174.

Garfield, E. (1955). Citation indexes for science. A new dimension in documentation through association of ideas. *Science*, *122*(3159), 108-111.

Goo, Kishore, Rao, & Nam. (2009). The role of service level agreements in relational management of information technology outsourcing: An empirical study. *MIS Quarterly*, *33*(1), 119-145.

Hui, Teo, & Lee. (2007). The Value of Privacy Assurance: An Exploratory Field Experiment. *MIS Quarterly*, *31*(1), 19-33.

Jones, B. F. (2010). Age and Great Invention. *Review of Economics and Statistics*, *92*(1), 1-14.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, *18*(1), 39-43.

Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book Company.

Kohlhase, M. (2010). *An open markup format for mathematical documents.* Available at http://www.omdoc.org/pubs/omdoc1.2.pdf

Koncel-Kedziorski, R., Bekal, D., Luan, Y., Lapata, M., & Hajishirzi, H. (2019). *Text generation from knowledge graphs with graph transformers.* arXiv. https://arxiv.org/abs/1904.02342.

Larsen, K. R., Hekler, E. B., Paul, M. J., & Gibson, B. S. (2020). Improving usability of social and behavioral sciences' evidence: A call to action for a national infrastructure project for mining our knowledge. *Communications of the Association for Information Systems*, *46*, 1-17.

Liu, A., Wu, Z., Michael, J., Suhr, A., West, P., Koller, A., Swayamdipta, S., Smith, N. A., & Choi, Y. (2023). *We're afraid language models aren't modeling ambiguity*. arXiv. https://arxiv.org/abs/2304.14399.

Liu, Y., Safavi, T., Dighe, A., & Koutra, D. (2018). Graph summarization methods and applications: A survey. *ACM Computing Surveys*, *51*(3), 1-34.

Locke, K., & Golden-Biddle, K. (1997). Constructing opportunities for contribution: Structuring intertextual coherence and problematizing in organizational studies. *Academy of Management Journal*, *40*(5), 1023-1062.

Matthews, D. (2021). Keeping up with the literature. *Nature*, *597*, 141-143.

Michels, C., & Schmoch, U. (2012). The growth of science and database coverage. *Scientometrics*, *93*(3), 831-846.

Musen, M. A. (2022). Without appropriate metadata, data-sharing mandates are pointless. *Nature*, *609*(7926), 222-222.

Negro, A. (2021). *Graph-powered machine learning*. Simon & Schuster.

Ou, C. X., Pavlou, P. A., & Davison, R. M. (2014). Swift Guanxi in Online Marketplaces: The Role of Computer-Mediated Communication Technologies. *MIS Quarterly*, *38*(1), 209-230.

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, *613*(7942), 138-144.

Pavlou, P. A., Huigang, L., & Yajiong, X. (2007). Understanding and Mitigating Uncertainty in Online Exchange Relationships: a Principal-Agent Perspective. *MIS Quarterly*, *31*(1), 105-136.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Peffers, K., Ein-Dor, P., & Zmud, R. W. (2003). Scholarly products in IS: Will advances in electronic media promote evolutionary or radical change? *The Communications of the Association for Information Systems*, *11*(1), 498-512.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280-291.

Piccoli, G., & Ives, B. (2003). Trust and the unintended effects of behavior control in virtual teams. *MIS Quarterly*, *27*(3), 365-395.

Rivard, S. (2020). Theory building is neither an art nor a science. It is a craft. *Journal of Information Technology*, *36*(3), 316-328.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379-423.

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). *Data mining for business analytics: concepts, techniques, and applications in R*. Wiley.

Song, Y., Watson, R. T., & Zhao, X. (2021). Literature reviewing: Addressing the jingle and jangle fallacies and jungle conundrum using graph theory and NLP. *Proceedings of the International Conference on Information Systems*.

Song, Y., Zhao, X., & Watson, R. T. (in press). Digitized knowledge-based literature reviewing: a tutorial on coding causal and process models as graphs. *Journal of Decision Systems*.

Sowa, J. F. (1999). *Knowledge representation: logical, philosophical and computational foundations*. Brooks/Cole.

Swanson, D. R. (1986). Undiscovered public knowledge. *The Library Quarterly*, *56*(2), 103-118.

Templier, M., & Paré, G. (2015). A Framework for Guiding and Evaluating Literature Reviews. *Communications of the Association for Information Systems*, *37*(6), 112-137.

Textor, J., Hardt, J., & Knüppel, S. (2011). DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology*, *22*(5), 745.

Thorndike, E. L. (1913). *An introduction to the theory of mental and social measurements*. Teacher's College, Columbia University.

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118-144.

Wagner, G., Lukyanenko, R., & Paré, G. (2021). Artificial intelligence and the conduct of literature reviews. *Journal of Information Technology*, *37*(2), 209-226.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: writing a literature review. *MIS Quarterly*, *26*(2), iii-xiii.

Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly*, *31*(1), 137-209.

Zhang, J. (2023). *Graph-ToolFormer: To empower LLMs with graph reasoning ability via prompt augmented by ChatGPT.* arXiv. https://arxiv.org/abs/2304.11116

Zhang, M., & Chen, Y. (2018). Link prediction based on graph neural networks. *Proceedings of the International Conference on Neural Information Processing Systems.*

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57-81.

Zupko, R. E. (1990). *Revolution in measurement: Western European weights and measures since the age of science.* American Philosophical Society.

About the Authors

Richard (Rick) Watson is the research director of Digital Frontier Partners and Regents Professor and J. Rex Fuqua Distinguished Chair for Internet Strategy Emeritus of the University of Georgia. He is a former president of AIS and was awarded its highest honor, a LEO. For about a decade, he was the research director for the Advanced Practices Council of the Society of Information Management and a visiting researcher at the Research Institutes of Sweden (RISE). In 2021, the University of Liechtenstein awarded him an honorary doctorate. He is an Honorary Visiting Professor at the Queensland University of Technology and a Schöller Senior Fellow at the University of Erlangen-Nuremberg. Professor Watson has published over 200 journal articles and written books on electronic commerce, data management, and energy informatics. His most recent book is *Capital, Systems, and Objects.*

Yuanyuan (April) Song is a fifth-year doctoral candidate in the Management Information Systems Department at the University of Georgia. Her research focuses on developing causal knowledge analytics methods and health information systems. She has been serving as a reviewer for conference proceedings, such as ICIS and AMCIS. She has published in *Journal of Decision Sciences* and in the AMCIS and ICIS proceedings.

Xia Zhao is an associate professor of management information systems at the University of Georgia. She received her PhD in management science and information systems from the University of Texas at Austin. Her research interests include the platform economy, IT governance, mobile computing, artificial intelligence, fintech, and knowledge analytics. She has published papers in *MIS Quarterly*, *Information Systems Research*, *Production and Operations Management*, *Journal of Management Information Systems*, *Decision Sciences*, *Decision Support Systems*, *Information Systems Frontier*, *IEEE Computer*, and *International Journal of Electronic Commerce*. She is a senior editor of *Decision Support Systems*.

Jane Webster received her PhD from New York University. She is the E. Marie Shantz Chair Emerita in Management Information Systems at Queen's University, Canada, and a Fellow of the *Association for Information Systems.* Jane served as a senior editor for *MIS Quarterly*, program chair for the International Conference on Information Systems, and as the VP of Publications for the Association for Information Systems. Most recently, she has studied ways in which to encourage more environmentally sustainable behaviors in organizations. Her research has also investigated the impacts of technologies in the support of distributed work, organizational communication, employee recruitment and selection, employee monitoring, training and learning, and human-computer interaction.