Deep Learning for Information Systems Research

Sagar Samtani, Hongyi Zhu, Balaji Padmanabhan, Yidong Chai, Hsinchun Chen, and Jay F. Nunamaker Jr.

© Kelley School of Business, Indiana University, Bloomington, IN; University of Texas at San Antonio, Alvarez College of Business, San Antonio, TX; Muma College of Business, University of South Florida, Tampa, FL; Hefei University of Technology, School of Management, Hefei, Anhui, China; Eller College of Management, University of Arizona, Tucson, AZ; Center for the Management of Information, Eller College of Management, University of Arizona, Tucson, AZ

ABSTRACT
Modern artificial intelligence (AI) is heavily reliant on deep learning (DL), an emerging class of algorithms that can automatically detect non-trivial patterns from petabytes of rapidly evolving “Big Data.” Although the information systems (IS) discipline has embraced DL, questions remain about DL’s interface with a domain and theory and DL contribution types. In this paper, we present a DL information systems research (DL-ISR) schematic that reviews DL while considering the role of the application environment and knowledge base, summarizes extant DL research in IS, a knowledge contribution framework (KCF) to position DL contributions, and ten guidelines to help IS scholars design, execute, and present DL for computational, behavioral, or economic IS research. We illustrate a research contribution to DL for cybersecurity. This article’s contribution to theory resides in the conceptual DL-ISR schematic and KCF, while its contributions to practice are based on its practical guidelines for executing DL-based projects.

KEYWORDS
deep learning; artificial intelligence; knowledge-contribution framework; information-systems methodologies; research guidelines; design science research; behavioral research; economics of IS

Introduction
Artificial intelligence (AI) has rapidly emerged as a key disruptive technology of the twenty-first century. Increasingly, many companies are capitalizing upon the initial promise of AI to conduct their business operations with unprecedented efficiency, effectiveness, and scale. Deep learning (DL) is the core engine that powers many modern AI innovations and development. In contrast to earlier AI approaches that relied on feature engineering or formal logic, DL can automatically extract insights from the ever-increasing data deluge and closely emulate a human’s thought processes and decision-making. These capabilities have helped DL quickly emerge as a viable solution for many high-impact applications and have led to considerable investments into developing novel research to solve grand societal challenges. Information systems (IS) researchers have embraced DL, and, more broadly, AI, as one of the critical areas for growth in the discipline, evidenced by the increased quantities of special issues, courses, and community events in the field.
Unlike traditional machine learning (ML) that made developments over 30-40 years, DL is progressing at a staggering rate on a day-to-day basis [41]. Increasingly, DL can operate on petabytes of rapidly evolving “Big Data” to automatically detect correlations and non-trivial patterns with minimal human intervention. Some may argue that these benefits alone warrant applying various DL models on datasets to produce and advance knowledge rather than following the well-established scientific method of developing theoretical assumptions and hypotheses followed by careful model development and experimentation [6]. However, irrespective of methodology or tradition, IS has historically been an application-driven, information-centric discipline that draws upon domain requirements from an application environment and theories or guiding principles from a knowledge base to guide studies. Applying DL models without careful consideration of these factors can result in an inability to define bounded studies, difficulty distinguishing between meaningful and spurious correlations, AI distrust, and an incomplete understanding of phenomena. Rapid iterative cycles of data-driven insights contributing to theorizing can drive progress in applied fields such as IS.

To date, IS scholars have provided detailed overviews for emerging methods, offered clear guidelines on how to conduct scientific research with the method, and/or conceptual frameworks for positioning contributions for the method in the areas of systems development methods [39], predictive analytics methods [57], interpretive research methods [26], grounded theory methods [60], methods for generalizability [31], and mixed-methods [59]. These contributions have led to substantial scholarly progress and practical impact. However, DL lacks similar guidelines and frameworks within IS. As a result, IS scholars can face several challenges when conducting DL research, including delineating how DL differs in IS versus related disciplines (e.g., computer science [CS]), identifying the role of theory, pinpointing how DL could be leveraged for multiple traditions of IS research (not just computational), balancing academic rigor and practical impact, and positioning their contributions. Therefore, the objectives of this article are as follows:

- First, we formulate a DL in IS research (DL-ISR) schematic that reviews DL fundamentals. In contrast to DL textbooks or academic tutorials that provide a technical, domain-agnostic summary of DL, the DL-ISR schematic presents a research-focused, abstracted view of the interplay between technical DL processes, application environment, knowledge base of theories and natural science inspirations, and business or societal outcomes.
- Second, to help illustrate how the DL-ISR schematic encompasses behavioral, economic, and computational IS traditions, we reviewed 14 IS journals to identify how IS scholars have used DL. This review also helps reveal potential future directions for IS research with DL.
- Third, we present a KCF to help IS scholars to clearly position their contributions to DL. This KCF aims to provide IS scholars with a mechanism to carefully articulate and position a wide array of DL-ISR contributions (e.g., computational, behavioral, economic) to help maximize their impact.
- Finally, we articulate ten guidelines to help IS scholars execute DL research (irrespective of IS tradition or paradigm), streamline their efforts, and provide a lens for reviewers and editors to evaluate DL research activities and contributions. These guidelines are based in design science research (DSR) principles and research
guidelines articulated by the IS discipline’s top journals. We summarize the major components of a rigorous DL evaluation as part of these guidelines. We illustrate a sample research contribution to DL within cybersecurity, position the contribution with the KCF, and summarize its alignment to the guidelines.

**Related Work**

Natural starting points for identifying how to use or study DL for behavioral, economic, or computational IS research are textbooks used in DL courses, academic survey papers, or academic tutorial papers. In Table 1, we present a review of selected recent conceptual textbooks, implementation textbooks, and academic surveys and tutorials focused on DL. For each resource, we reviewed the domain, discipline, or application area the authors situate DL, and their DL algorithm coverage. We also identified if each resource provides research guidelines (denoted as RG) on how to execute research or a contribution framework (denoted as CF) for the audience to consider when developing their DL approaches or studying phenomena related to DL.

Conceptual textbooks often provide a mathematical or theoretical description of core DL architectures (e.g., artificial neural network [ANN], convolutional neural network [CNN], recurrent neural network [RNN]) and learning strategies (e.g., deep reinforcement learning [DRL], transfer learning [TL]) [4, 11, 21, 29]. Implementation-focused textbooks provide code examples and technical illustrations on how to apply existing DL algorithms on publicly accessible datasets (e.g., MNIST for image recognition). Most conceptual and

| DL RESOURCE CATEGORYa | Year | Author | Domain, discipline, or application area | DL algorithm coverage | RG? | CF? |
|------------------------|------|--------|----------------------------------------|-----------------------|-----|-----|
| Conceptual Textbooks   | 2020 | Calin [11] | Agnostic | ANN, CNN, RNN, GAN | No | No |
|                        | 2020 | Krohn et al. [29] | Agnostic | ANN, CNN, DRL, TL | No | No |
|                        | 2018 | Aggarwal [4] | Agnostic | DNN, RMB, RNN, CNN, DRL | No | No |
|                        | 2016 | Goodfellow et al. [21] | Agnostic | DNN, AE, CNN, GDM | No | No |
| Implementation Textbooks | 2022 | Raff [42] | Agnostic | ANN, CNN, RNN, CL, TL | No | No |
|                        | 2021 | Glassner [20] | Agnostic | ANN, CNN, AE, RNN, BERT, Attention, DRL, TL, GAN | No | No |
| Academic Surveys and Tutorials | 2021 | Ramsundar et al. [52] | Life sciences | RNN, VAE, GAN | No | No |
|                        | 2019 | Foster [19] | Art, composition | GAN, CNN, AE, VAE, BERT | No | No |
|                        | 2022 | Cao et al. [12] | Finance | DNN, RNN, CNN, DRL, GNN | No | Yes |
|                        | 2021 | Minaee et al. [37] | Image analysis | CNN | No | Yes |
|                        | 2021 | Bengio et al. [7] | Agnostic | UL, SSL, CL, VAE | No | No |
|                        | 2021 | Abukmeil et al. [2] | Agnostic | GAN, UL | No | No |
|                        | 2021 | Chen et al. [69] | AR | RNN, CNN, MVL | No | Yes |
|                        | 2019 | Zhang et al. [64] | Recommender Systems | DNN, AE, CNN, RNN, RBM, GAN, RL, Attention | No | Yes |
|                        | 2018 | Cai et al. [10] | Agnostic | Graph embedding techniques | No | Yes |
|                        | 2015 | LeCun et al. [30] | Agnostic | CNN, RNN, LSTM | No | No |

Note: aThis review excludes online non-peer reviewed tutorials (e.g., from Medium, Towards Data Science, Machine Learning Mastery). These tutorials have the same limitations as it pertains to DL-ISR as the listed resources. AE, autoencoder; ANN, artificial neural network; AR, activity recognition; BERT, bidirectional encoder representations from transformers; CF, contribution framework; CL, contrastive learning; CNN, convolutional neural network; DGM, deep generative models; DNN, deep neural network; DRL, deep reinforcement learning; GAN, generative adversarial networks; GNN, Graph Neural Network; MVL, multi-view learning; LSTM, long-short term memory; RG, research guidelines; RNN, recurrent neural network; RBM, restricted boltzmann machine; SSL, self-supervised learning; TL, transfer learning; UL, unsupervised learning; VAE, variational autoencoder.
implementation textbooks are targeted at industry professionals looking to implement routine designs of DL, agnostic of any single domain or application area, and focus on providing instruction on common DL algorithms. As such, these resources do not provide guidelines about how to execute DL for research applications. Similarly, these books are often devoid of perspectives on the types of research contributions researchers can make. Academic surveys and tutorials often provide a description of how DL has been used in various domains, disciplines, or application areas, including finance [12] and image segmentation [37]. While these papers often provide an excellent, focused overview of the state-of-the-art practices of DL, they do not provide explicit guidelines on how to execute DL-based research. Some works provide a summary of promising future directions of DL that researchers could make contributions to [10, 12, 37, 64]. However, these summaries are often focused on how to extend existing algorithms in a manner that could enhance the performance of well-defined applications.

All categories of DL resources lack a description about how economic or behavioral studies could leverage DL and the role of theory in DL design and evaluation of DL-enabled systems (e.g., studying issues of AI and DL trust, bias, fairness and ethics, DL adoption, use, and impact, etc.). While IS scholars have largely embraced AI, extant reviews on the subject have often focused on the phenomena around DL (omits computational research) and often focus on a singular aspect of DL or AI, such as the role of general ML in the developing world [14], algorithmic bias [28], human-AI interfaces [25, 35], and managing AI [8]. Consequently, there remains a significant need for an abstracted lens (micro-macro view) of how the technical DL processes (micro-view) is situated within IS discipline (macro-view) in a manner that considers multiple IS traditions and research paradigms, a perspective on the types of contributions IS scholars can make to DL, and guidelines that IS scholars can consider when developing and positioning their DL research.

**DL-ISR**

Past IS literature introducing a method to the IS discipline (an applied discipline) have typically reviewed the fundamentals of the method. DL is a method that can offer opportunities to multiple IS traditions, and therefore needs to be reviewed in a manner that considers the role of the application environment (where domain problems are identified), knowledge base (that informs the research execution), and business or societal outcomes. DL has several unique characteristics compared to existing IS research methods, AI-based data analytics techniques (e.g., ML), or other emerging technologies (e.g., blockchain) that requires a specialized review and reflection. First, standard ML algorithms are trained on feature vectors that are extracted (often based on considerable domain knowledge and manual efforts) from raw data. In contrast, DL algorithms are based in connectionist principles from cognitive science and can therefore automatically learn feature representations at different levels of abstraction from raw data in their natural form (e.g., raw sensor data, images, text, etc.). As a result, DL can attain significantly higher performances than traditional ML algorithms, excel in dynamic scenarios where feature sets are often evolving (e.g., cybersecurity), and can generalize across multiple applications. However, DL algorithms are unique in that they are “black-boxes” compared to their ML counterparts (e.g., decision tree), replacing or augmenting decision-making processes in high-impact application environments (e.g., medicine, aerospace, military), being deployed “at-scale” (e.g.,
Netflix recommendation system), and revealing deep-seated biases within society. These issues have brought into question significant issues of AI trust, fairness, and transparency and has caused new questions to arise about AI system development, adoption, usage, governance, and maintenance. These issues significantly alter how DL influences business or societal outcomes. Other emerging technologies, such as blockchain or internet-of-things (IoT) systems lack the same data-centricity and analytics capabilities of DL models and are largely devoid of the aforementioned concerns.

With DL’s unique characteristics in mind as well as the limitations of extant DL reviews, we present a DL-ISR schematic in Figure 1. We use the term schematic as it has been leveraged by past IS scholars summarizing the fundamentals and steps of executing a method (predictive analytics) when presenting the technical components of a method within the larger IS discipline [39, 57]. Five major components comprise the DL-ISR schematic: (1) application environment, (2) knowledge base, (3) standard ML processes, (4) deep learning design, and (5) business and societal outcomes. The core technical DL component is color-coded in gray.

The application environment comprises people, organizations, and technologies, and defines the relevant issues for study. The knowledge base offers IS scholars theories (e.g., computational, social-behavioral, and economic (SBE), design) and natural science inspirations to draw upon to execute their DL research. Similarly, the knowledge base is where the research contributions from past studies reside. Standard ML processes include data, tasks, and requirements. The DL design component encompasses encoding, learning paradigms, architecture, learning strategy and implementation, and model outcome. Each component helps drive business and societal outcomes, including use, adoption, continuance, privacy, business impact, bias, fairness, and ethics (all unique in the context of DL). Concepts within and across components are linked through a series of relationships. The semantics of the

![Figure 1](image-url)
relationships between concepts are summarized in Table 2 and in the following subsections. The bolded, italicized text represents the logical flow of activities researchers often follow. To make the description tangible, we provide a running example of how each component, concept, and relationship is instantiated using the major success made by Google DeepMind that developed a novel DL approach for early acute kidney injury (AKI) alerting by examining the knowledge embedded in electronic health records (EHR) [58].

Application Environment

It is generally well known in IS that the Application Environment comprise people, organizations, technologies, and governance [24]. As such, each component of the environment defines relevant issues for the domain question or problem being studied and guides, grounds, or informs the examination of the business or societal outcomes of a DL-based approach. Grounding research inquiry within key environmental issues keeps the problem space bounded [39]. In the AKI detection paper, the problem is motivated by AKI’s high prevalence in U.S. inpatients (1 in 5), the need to reduce the high death rates associated with related admissions, and the challenge of providing clinicians with intelligent, automatic, real-time, and personalized AKI.

Knowledge Base

While the application Environment helps define relevant issues for the domain problem, the knowledge base serves as a resource for researchers to draw appropriate knowledge to develop and refine their DL research studies [22]. Since DL is a method that holds value for multiple IS paradigms, the knowledge base includes four types of knowledge [7, 21, 24, 30]: (1) Computational theories such as information theory, graph theory, or statistical learning theory, (2) SBE Theories such as game theory, (3) design theories for designing artificial systems, and (4) natural science inspirations from biology, neuroscience, physics, and others. A technical example of how the knowledge base can inform DL processes is game theory for adversarial learning [21]. From a behavioral perspective, the knowledge base can also provide theories that can help design studies examining the effects of DL systems (e.g., AI trust, bias). For example, human-computer interaction (HCI) theories can guide studies on how user interfaces (UIs) for DL-enabled chatbots are adopted. The Knowledge Base also includes knowledge developed from a study, including design theory, the role of theory in particular domains, and where specific knowledge is abstracted to general insights. DL models developed as part of a researcher’s activities (e.g., a custom RNN for codifying qualitative data) can therefore be part of the knowledge base.

Standard Machine Learning (ML) Processes

Since DL is a branch of ML, fundamental ML principles serve as the foundation for successful DL projects. Within DL-ISR, standard ML processes begin with a domain problem that is defined based on relevant issues within the application Environment. This domain problem guides scholars in three key activities. First, the domain problem helps identify the relevant data within the environment. These data can be generated from multiple sources, including humans (e.g., text, sound, video), machines (e.g., log files), or
Table 2. Semantics of relationships between major DL-ISR components.

| Component | Semantics of the relationship between concepts | Running example from Tomašev et al. [58] |
|-----------|-----------------------------------------------|----------------------------------------|
| Application Environment and Knowledge Base | Application environment and knowledge base defines relevant issues or has appropriate knowledge drawn, respectively, for the problem | • The need to reduce clinician workloads and improve patient outcomes drives the need for accurate and timely AKI alerts. |
| Standard ML Processes | Domain problem helps identify required data | • The knowledge drawn from the knowledge base included RNN, multi-task learning, and deep residual embeddings. |
| Deep Learning Design | Data requires an encoding | • While the DL engine provides AKI alerts, how this is embedded within a new treatment protocol will affect improvements in both patient outcomes and overall costs. |
| Business and/or societal outcomes | Learning Strategy and Implementation produces the Model outcome | • The problem requires a ground truth AKI dataset. |
| | Business and Societal Outcomes enhances or improves the application environment and contributes new knowledge to the Knowledge Base | • The tasks are formulated as predictive (AKI onset, laboratory testing results). |

Notes: Bolded, italicized text represents the logical flow of activities researchers often follow. AKI, acute kidney injury; EHR, electronic health records; RNN, recurrent neural network; DL, deep learning; ML, machine learning.

a combination of both [13]. Second, the domain problem helps formulate the predictive and explanatory ML tasks. Predictive tasks train algorithms to execute a pre-specified goal for enhanced decision-making and prescriptive applications [57]. Explanatory tasks aim to describe and explore data via unsupervised algorithms, summary statistics, visualizations, correlations, or rules [1]. Finally, the domain problem helps specify requirements essential for the environment. These can range from industry-specific regulations, organization-specific processes, and end-user needs (interpretability). The emphasis on data identification, task formulation, requirements specification, and instantiation of a DL process can vary greatly based on their breadth, depth, evolution, and other factors. Within the
healthcare example, the data were carefully divided into broad segments to help increase the generalizability of the researchers’ approach and support the requirements of early prediction on the scale of days.

Deep Learning Design

DL Design can be decomposed into encoding, learning paradigms, deep architecture, learning strategy and model implementation, and model outcome [30]. Each is detailed in turn.

Encoding

Data requires pre-processing and encoding to produce a structure recognizable by prevailing DL architectures. Encoding is the process of converting the raw data into a specified format that preserves particular types of information [21]. For example, sequences can capture local and global temporal dependencies in text. Grids can encompass images, IoT device signals, temporal sequences (e.g., stock prices), and spatial data. Data sources can have multiple modalities, representations, and encodings simultaneously. Examples include video (spatial-temporal), attributed social networks (text, image, graph), and others. Encoding selection varies based on available data, key structural properties researchers wish to capture for their application, and/or relevant kernel theories. Within the AKI example, the encoding was developed by organizing EHR data sequentially based on time for their predictive analytics task.

Learning Paradigms

There are numerous ML paradigms beyond the traditional predictive and explanatory dichotomy. For example, one could go with a traditional supervised learning paradigm or transfer learning paradigm [34, 61], which is built upon the concept that an existing model can be reused intelligently with new data. Within DL-ISR, the combination of data, tasks, and requirements determines how the learning paradigm is operationalized. For example, the data characteristics (e.g., richness of labels, volume) dictates the selection of semi-supervised, unsupervised, multi-view, multi-source, contrastive, and online learning [7]. Precise task specifications such as handling multiple datasets simultaneously, transferring knowledge across domain tasks, rewarding correct algorithmic behavior, and others, can help in selecting a finer-grained learning paradigm such as multi-task [66], transfer, reinforcement [38], and variational inference [63]. Finally, end-user requirements identified from the Application Environment such as stability, privacy preservation, real-time processing, human involvement, interpretability or theoretical foundations drawn from the knowledge base can motivate the selection of adversarial [21], federated [62], active [16], and/or interpretable [17] learning strategies. In the AKI context, the predictive multi-task model captures and models multiple inputs simultaneously. The loss function consists of AKI prediction loss and seven lab-test predictions.

Deep Architecture

Encodings such as sequences, grids, and graphs require model designs (i.e., architectures) with the capacity to transform and manipulate such structures. Deep architectures generally comprise three building blocks: primitives, basic processing units, and extensions.
Primitives such as neurons, activation functions, layers, weights, and error correction can be organized into functioning groups to form basic processing (i.e., neural) units such as the ANN for feature vectors, CNN for grid-structured data, RNN for sequential data, and Graph Neural Network (GNN) for network structures. Extensions to these units increase their capacity to learn. Common extensions include capturing forward and backward processing for the bi-directional long short-term memory (BiLSTM), attention mechanism \cite{17, 49} integration for enhanced model interpretability, and others. Architecture selection is guided by the learning paradigm and data encoding. The development of an architecture can be inspired from the knowledge base. For example, the convolutional and pooling layers in the CNN were inspired by the processes of the human visual cortex \cite{30}. Each basic processing unit can be extended to account for additional data characteristics, task specifications, and environmental requirements. When considering the AKI example, the model was an LSTM customized to incorporate key EHR data characteristics.

**Learning Strategy and Model Implementation**

Selecting a strategy on how and what the model learns follows after selecting the data encoding, learning paradigm, and deep architecture. Analogous to model estimation in conventional ML, the learning strategy is how the model is trained to execute the task. Strategies can focus on maximizing performance with loss functions, optimizing gradient descent through batch normalization, dropout, joint training, generalizing or regularizing the architecture with noise injections, parameter restrictions, and specifying activation functions and number of layers to improve a model’s performance. The learning strategy and overall DL process are implemented into hardware and software.

**Model Outcome**

Results from a DL procedure can oftentimes be the conclusion of an analytics-driven project or serve as input for a subsequent process. The former often manifests itself in the form of predictive analytics (e.g., prediction of AKI). However, DL outputs can also be a final or intermediate data representation (e.g., embedding, independent variable), pattern, synthetic data, or something else. Each can be integrated into subsequent (i.e., downstream) post-processing tasks, larger workflows (e.g., RapidMiner style), econometric models, and more. The model outcome is the final outcome at the end of a complete technical workflow. Figure 1 captures this by noting that the core DL engine components help produce the model outcome and are not necessarily the model outcome themselves. However, they can serve as such, if there are no other workflows.

**Business and/or Societal Outcomes**

Attaining strong model outcomes (e.g., performance, quality) does not necessarily lead to the optimal business outcomes \cite{23}. Achieving maximal business and/or societal outcomes is influenced by how factors from the application environment and the model outcome interface. Common business and/or societal outcomes that can be examined include measures of business impact, privacy, bias, and fairness. Ultimately, the outputs of the study of Business and Societal Outcomes for both technical and non-technical (behavioral and economic studies) should enhance or improve application environment and contribute new knowledge to the knowledge base.
Methodology

Articles in top IS journals that present methods for IS research typically review how the method has been used in publications at top IS journals, provide a summary (or framework) of the types of contributions researchers can make with the method, and/or summarize guidelines IS scholars can consider when developing their papers [26, 31, 39, 57, 59, 60]. Therefore, we executed three sets of activities: (1) A review of extant IS literature pertaining to DL, (2) development of a KCF, and (3) an articulation of guidelines for executing DL-ISR. We describe the method for executing each activity in the ensuing sub-sections.

A Review of Extant IS Literature Pertaining to DL

IS literature presenting a method often review prevailing IS journals to identify the extent (i.e., current status) the method has been used [57, 59]. Such a review can help identify contributions made by IS scholars, develop guidelines on how the method could be leveraged in future IS research, and/or point to future directions. DL is a method that has the potential to be used in behavioral, economic, and computational IS research. Therefore, a goal of this review was to examine studies that developed novel DL designs or contributed to our understanding of DL through behavioral and economic methods. Since IS scholars are often evaluated for promotion and tenure based on journal publications, we selected IS journals in the “IS Senior Scholars Basket of Eight,” Financial Times (FT) 50 List, UTD 24, and if they are well-regarded by IS scholars for accepting DL-related work. 14 journals were selected for examination. In alphabetical order, the journals selected were: ACM Transactions on MIS (TMIS), Communications of the Association of IS (CAIS), Decision Support Systems (DSS), European Journal on IS (EJIS), INFORMS Journal on Computing (JoC), IS Journal (ISJ), IS Research (ISR), Journal of the AIS (JAIS), Journal of Information Technology (JIT), Journal of Management Information Systems (JMIS), Journal of Strategic IS (JSIS), Management Science (MS), MIS Quarterly (MISQ), and MISQ Executive (MISQE).

We used the ACM Digital Library, AIS eLibrary, EBSCOHost, INFORMS PubsOnLine, and Science Direct to search for articles. To identify relevant articles, we searched the title, abstract, and keywords of each paper for the keywords of “deep learning,” “artificial neural networks,” “neural networks,” “neural,” “artificial intelligence,” “embed,” and “embedding.” We executed our search over the last five full years of publications (2016-2021), as this time period represents DL’s rapid proliferation. We omitted editorials, commentaries, opinions, debates, perspectives, and viewpoints, as these papers only offer non-empirical opinions or conclusions about AI or DL. The quantities of included papers were tabulated on a year-by-year basis.

After retrieving and filtering papers, we downloaded the full text of the articles for review and coding. Since the DL-ISR schematic reviewed DL in a manner that encompasses behavioral, economic, and design science IS traditions, we asked each coder to examine the motivation, context, research methods, and implications specified in these papers and code the contents of the papers to five major components (and their sub-components) of the schematic: (1) application environment, (2) knowledge base, (3) standard ML processes, (4) DL design, and (5) business and societal outcomes. Mapping retrieved papers to the review of the method’s overview (in this article, DL-ISR schematic) is a commonly accepted practice when illustrating how a method has been used in the IS community [31, 59]. This
process was executed by two coders, each with nearly a decade of executing DL research. We computed the inter-coder reliability (based on Cohen’s Kappa) to ensure consistency of coding. Each coder rated the papers separately (i.e., independently) over two weeks. The Cohen’s Kappa score was computed after the first round of coding. The two coders met after the first round of coding to discuss and resolve minor discrepancies until 100 percent agreement was achieved. After the coding was completed, we grouped the papers into: (1) studies focused on DL design (i.e., used or extend a deep architecture) and (2) studies that examine DL outcomes (i.e., with no technical DL design). Segmenting articles in this fashion helps provide an understanding of how technical, behavioral and economic methods have been used to study DL phenomena.

**Development of a KCF**

DL is an area where multiple disciplines can make research contributions. However, these contributions can be different in nature. For example, CS scholars often operate in large labs and have close collaborations with technology giants such as Google, Facebook, Tesla, Twitter, Netflix, and Apple. Such entities often have large datasets, computing resources, and engineering teams to facilitate fundamental DL research contributions (e.g., large pre-trained models, DL paradigms). IS may lack deep connections with technology giants but often have excellent inter-disciplinary relationships with corporate partners ranging from start-ups to Fortune 100 companies that can inspire computational, behavioral, or economic DL research [45]. Moreover, IS often serves as the technical interface to other business school disciplines (e.g., economics, operations management, supply chain management) and vertical application areas (e.g., healthcare, retail, e-commerce, energy, transportation). However, there currently lacks a clear, specialized perspective on the types of DL-oriented contributions IS scholars can make. Grouping contributions into major categories can help authors, reviewers, and editors quickly position and ascertain the novelty of DL-based research and accelerate DL-ISR.

Past IS scholars have provided framework for positioning contributions when there is a potential to use the method directly to generate research insights or contribute directly to the method (e.g., extend for a particular application) [26,31,60]. Since DL is a method that can be extended (e.g., to make technical contributions), used directly to enhance existing methods (e.g., econometrics, qualitative research), or be treated as a black-box IT artifact (technology, tool) that can produce outcomes for study (e.g., bias, fairness, adoption), we propose a KCF for positioning DL research in IS. Past IS scholars have developed their frameworks for positioning contributions based on existing frameworks [60], characteristics of the method being examined [31], or the author’s own experiences with the method [26]. Rather than relying on our own experience alone, we aimed to ground the proposed KCF in the CRoss-Industry Standard Process for Data Mining (CRISP-DM) to help ensure the proposed guidelines are inclusive for multiple paradigms of IS research. The CRISP-DM framework comprises: (1) Problem or business understanding that focuses on formulating an ML or DL solution based on the core business environment’s needs or requirements, (2) Data understanding and data preparation to understand and construct representations for modeling, (3) Modeling and evaluation that focuses on developing a model and evaluating its technical performance, and (5) Deployment that examines how well the model performs in the business environment it was designed for. Although not specifically focused on
DL-oriented research, CRISP-DM is an ideal framework to underpin the KCF for two key reasons. First, in contrast to data mining frameworks that present only technical activities (e.g., knowledge discovery in databases), CRISP-DM places a significant focus on carefully understanding the business (application) environment when developing a model and examining the model’s outcomes when deployed. The focus on business understanding, deployment outcomes, and replicability of processes is consistent with the manner we presented the DL-ISR schematic. Second, many specialized data mining processes and data science trajectories for high-impact applications environments (e.g., tourism management, insurance refinement, payment management) are rooted in CRISP-DM [36].

**An Articulation of Guidelines for Executing DL-ISR**

The proposed KCF can help provide a unified lens for articulating and evaluating DL research contributions. IS scholars have indicated that providing a set of strategies, guidelines, or principles for authors to consider when developing their research. These guidelines are essential for authors to develop, refine, and strengthen their manuscripts and for reviewers and editors to systematically evaluate manuscripts to make informed decisions. Guidelines are typically developed based on the author’s own experiences with the method [26, 59], based on an extant framework [26, 39, 57], or a combination thereof [26, 39]. Therefore, guidelines also typically encompass practical and operational considerations (e.g., data collection strategies, method validation, etc.).

A natural starting point for IS scholars to seek guidance on how to execute DL research is the design science research (DSR) paradigm. DSR is a paradigm that is synergistic with other paradigms of IS research (e.g., behavioral, economic). DSR offers seven research guidelines: (1) design as an artifact, (2) problem relevance, (3) design evaluation, (4) research contributions, (5) research rigor, (6) design as a search process, and (7) communication of research. Thanks to the breadth of applications and methods in the IS discipline, four genres of DSR have emerged [43]: (1) computational that focuses on computational algorithms, (2) optimization that solve decisional issues, (3) representation that aims to represent phenomena, and (4) economics that looks to design mechanisms. Each genre has unique characteristics that require specialized guidelines that go beyond yet align with the original DSR principles and prevailing perspectives for scientific research articulated in top journals. Some of DL’s unique characteristics include specialized infrastructure requirements (e.g., GPUs) that necessitate the disclosure of model parameters, rapid design cycles and variety of designs due to DL’s numerous components that can be varied innumerably, and DL’s potential to be used in economic or behavioral research. Taken together, these unique characteristics necessitate specialized guidelines.

**Results of Review of Extant IS Literature Pertaining To DL**

The results of our initial search resulted in 269 articles across the 14 selected journals. 197 articles remained after applying the aforementioned filtering criteria. This quantity represented 4.101 percent of all of the articles published in the reviewed journals during the 2017-2021 timeframe. In Table 3, we present a count of the number of articles per year for each journal.
Table 3. Number of DL-related papers based on journals from 2017-2021.

| Journal (listed alphabetically) | 2017 | 2018 | 2019 | 2020 | 2021 | Total |
|---------------------------------|------|------|------|------|------|-------|
| ACM TMIS                        | 1    | 1    | 2    | 10   | 20   | 34    |
| CAIS                            | 0    | 0    | 1    | 1    | 1    | 3     |
| DSS                             | 7    | 13   | 8    | 19   | 29   | 76    |
| EJIS                            | 0    | 1    | 0    | 0    | 1    | 2     |
| INFORMS JoC                     | 0    | 0    | 0    | 1    | 5    | 6     |
| ISJ                             | 0    | 0    | 0    | 0    | 0    | 0     |
| ISR                             | 0    | 1    | 0    | 4    | 13   | 18    |
| JAIS                            | 0    | 0    | 1    | 1    | 5    | 7     |
| JIT                             | 0    | 0    | 0    | 0    | 3    | 3     |
| JMIS                            | 0    | 4    | 0    | 6    | 5    | 15    |
| JSIS                            | 0    | 0    | 0    | 1    | 0    | 1     |
| Management Science              | 0    | 0    | 3    | 0    | 2    | 5     |
| MIS Quarterly                   | 2    | 1    | 0    | 4    | 13   | 20    |
| MIS Quarterly Executive         | 0    | 1    | 1    | 4    | 1    | 7     |
| Total                           | 10   | 22   | 16   | 51   | 98   | 197   |

DL, Deep Learning; ACM TMIS, ACM Transactions on MIS; CAIS, Communications of the Association of IS; DSS, Decision Support Systems; EJIS, European Journal on IS; INFORMS JoC, INFORMS Journal on Computing; ISJ, IS Journal; ISR, IS Research; JAIS, Journal of the AIS; JIT, Journal of Information Technology; JMIS, Journal of Management Information Systems; JSIS, Journal of Strategic IS; MIS, Management Information Systems.

The quantity of DL-related papers for each journal consistently increased over the five-year period (from 10 in 2017 to 98 in 2021). DSS had the most DL-related papers, with a total amount of 76. ACM TMIS (34), MISQ (20), ISR (18), and JMIS (15) were the journals with the next four. The remainder of the reviewed journals (CAIS, EJIS, JoC, JIT, JAIS, MS, MISQE, and JSIS) had less than 10 articles each. After the first round of coding, the inter-rater reliability between our raters was .835. The coders met to resolve differences until the coding agreement was 100 percent. Due to space limitations, we omitted the fully coded taxonomy and all references included in the review from the main text and instead provide a summary of key takeaways. Readers can contact the lead author of this article to request the full taxonomy. Below we provide a summary of papers focused on technical DL design and a summary of papers studying the outcomes of a DL-based system.

Summary of Technical DL Design Studies

A total of 146 of the 197 (74.11 percent) reviewed papers leveraged or extended an existing DL architecture. 65 of these papers were published in DSS, 32 in ACM TMIS, 12 in JMIS, 12 in MISQ, 10 in ISR, six in INFORMS JoC, four in JAIS, two each in CAIS and MS, and one in EJIS. The top five most common domains of study were healthcare (31 papers), BI (26 papers), e-commerce (23 papers), cybersecurity (13 papers), and finance (six papers). Other domains included fashion, marketing, and gig economy. The primary business or societal outcome pertained to improving an organizational process (e.g., improving strategic decision making) or enhancing existing technologies (e.g., Covid-19 diagnosis tools, malware detectors). A total of 24 of 146 (16.44 percent) studies used an SBE related theory (e.g., decision, regret, affect) to guide their DL study. The rest drew upon knowledge about existing DL architectures and designs from the knowledge base to guide their DL design. 73 studies employed the core ANN architecture for their analysis in a supervised (predictive) setting using conventional feature vectors as encodings. CNN (33 studies) operating on grids and RNN (40 studies) operating on sequential data were the next most popular
architectures. Irrespective of architecture, only 27 studies extended the base model, with the most common extensions being the incorporation of bidirectional processing or attention mechanisms.

**Summary of Studies Examining DL Outcomes**

A total of 51 of the 197 (25.89 percent) of the reviewed studies did not implement any basic DL model architecture, but instead studied the outcome of an existing DL-based approach or system. When considering the application environment, 20 studies examined organizations, 15 studied technologies, 13 studied people, and three studied governance (particularly governance of algorithmic decisions). 15 studies used DL approaches (primarily word2vec) to produce an embedding for input into an econometric model. 36 studies leveraged behavioral techniques (primarily surveys and case studies) to examine phenomena related to DL-based systems, or more broadly, AI-enabled systems. Common areas of study included human-machine augmentation, AI-based hiring, investment in AI, and AI management. Similar to the DL-design studies, the most common business or societal outcome focused on business impact, particularly around enhancing existing processes or identifying the impact of AI within an organization. However, unlike the DL-design studies, studies examining DL outcomes also focused on topics such as bias, fairness, and ethics around DL and AI-enabled systems. None of the studies in our review leveraged DL for enhancing behavioral methods.

**KCF for DL-ISR**

There is significant potential for IS scholars to develop novel DL designs for emerging application areas and to enhance methods commonly used in economic or behavioral research such as generating variables for econometric analysis, automatically synthesizing text, sound, and image data for qualitative and mixed-methods research, coding semi-structured interviews, and more. Similarly, there remains opportunity for IS scholars to study the adoption, bias, etc. of DL-based systems on a particular application environment to draw insights about DL management, governance, strategic investment, organization processes, and regulations. Given the scope of potential future contributions, we developed a KCF for DL based in the well-established CRISP-DM with five major contribution types: (1) domain-specific, (2) representation, (3) learning, (4) system, framework, or workflow, and (5) innovation-Accelerating. Contributions can solely reside in a single type or span multiple types. In Table 4, we summarize each contribution type, their CRISP-DM grounding, definitions, map each contribution type to the DL-ISR schematic, and provide references from IS scholars. In the following subsections, we describe each contribution type and summarize common pitfalls IS scholars may face when aiming to produce a contribution.

**DL-ISR Contribution Type 1: Domain-Specific Novelty**

Since IS scholars excel in asking innovative, application-driven research questions, the first major type of DL-ISR contribution scholars can make is to formulate and execute DL processes in a new application area. Two major categories of domain-level contributions
Table 4. Proposed knowledge contribution framework (KCF) for DL-ISR.

| CRISP-DM grounding       | Contribution type     | Definition(s)                                                                 | Mapping DL-ISR schematic                                                                 | References from IS scholars |
|--------------------------|-----------------------|------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|------------------------------|
| Business Understanding   | Domain-specific       | ● Formulating and executing DL in a new application domain or behavioral or economic method  
                            | ● How the environment influences business outcomes given DL                     | Application Environment  
                            |                                                                              | Kokkodis [27]; Zhu et al. [67, 68]                                        |
| Data Understanding       | Representation        | ● Representing raw data in a novel structure for DL based on SBE theories; or a new architecture of the underlying model for DL | Knowledge Base  
                            |                                                                              | Kokkodis [27]; Zhu et al. [67, 68]                                        |
| and Data Preparation     | Learning              | ● Developing new techniques for learning/estimating DL models or developing novel specifications of the important components | Learning strategy and implementation  
                            |                                                                              | Ahmad et al. [5]                                                            |
| Modeling and Evaluation  | System, Framework, or | ● Creating a new way of integrating DL in a broader system  
                            |                                                                              | Knowledge Base  
                            | Workflow Innovation-Accelerating                                           | Adamopoulos et al. [3]; Kokkodis [27]; Liu et al. [32]; Liu et al. [33]; Schuetzler et al. [56]; Zhang and Ram [65] |
|                          | System, Framework, or | ● Studying behavioral or economic issues around DL systems  
                            |                                                                              | Model outcome  
                            | Workflow Innovation-Accelerating                                           | Samtani et al. [55]; Zhu et al. [67]                                      |
|                          | Workflow Innovation   | ● Developing mechanisms to enhance a study’s reproducibility; sharing code and data | Application Environment  
                            |                                                                              |                                                                              |                                                                              |

Notes: DL-ISR, deep learning-information systems research.
can be made. The first pertains to applying emerging DL architectures and/or learning paradigms in a new, high-impact domain that has traditionally leveraged conventional ML or non-automated analyses. Such applications can yield significant performance gain and result in high-impact outputs (e.g., AKI research) and facilitate behavioral studies that examine how the environment (e.g., theories, processes, organizations, etc.) interacts with DL-based model outcomes. The second type of contribution focuses on formulating new questions within a domain that were previously non-trivial without DL. For example, DL could prove valuable for researchers engaging in qualitative or mixed-methods research to effectively process unstructured data collected from interviews or produce meta-inferences. IS scholars engaging in economics of IS research may consider using DL models in their workflow encode for complex data structures (e.g., spatial-temporal graphs) for input into econometric models.

There can be three key pitfalls in attaining domain-level novelty. First, contributions can be marginal, incremental, and/or obvious when the domain is well-studied or does not require DL. In the situations where DL design is not the focus, the study should still be executed in a manner consistent with the best practices in that paradigm. Second, incorrect DL processes can be selected when scholars acquire new datasets and directly apply well-established DL techniques without carefully considering the data characteristics, tasks, or domain requirements. Finally, contributions can often stay at the technical discussion (e.g., DL outperforms standard ML) without considering the value of the proposed processes to the domain. Work exhibiting these issues are commonly rejected at top IS journals [23]. Each issue can be mitigated by articulating the importance of the domain, justifying why DL is necessary, demonstrating non-trivial technical improvements, and attaining non-obvious domain improvements of value to relevant stakeholders.

**DL-ISR Contribution Type 2: Representation Novelty**

DL’s success is often contingent upon how raw data from an environment are encoded. These encodings help motivate carefully constructed architectures that can fully capture the encoding’s semantics and the environment’s underlying data-generating processes. Studies that design, develop, and evaluate novel encodings or architectures based on domain considerations, computational theory, and/or relevant SBE perspectives are representation-level contributions. Leveraging SBE theories to guide the selection of feature representations or using behavioral approaches to evaluate the encodings learned from technical DL processes (e.g., human-AI interfaces, augmented intelligence) are natural synergies between behavioral and technical DL research. Improving a model’s capacity to learn is often the key motivation behind the development of novel architectures. Examples of architectural contributions include carefully constructed attention mechanisms to enhance model interpretability, new processing units, and enhancing models with additional processing capacity.

Common challenges for researchers aspiring to attain representation-level novelty include insufficient grounding in domain requirements, data characteristics, or theory. The novelty offered may also be trivial, incremental, or marginal. A common example is concatenating features to embeddings. These issues can be proactively mitigated by clearly delineating, justifying, and demonstrating the performance benefits between the proposed
encoding and the original. A key issue pertaining to the architecture’s novelty is claiming that common engineering (e.g., stacking layers) considerations that are often varied during sensitivity or ablation analyses are new.

**DL-ISR Contribution Type 3: Learning Novelty**

A key reason for DL’s rapid growth has been its ability to automatically learn (i.e., estimate and fit parameters) from encodings and architectures. The learning-level novelty in the KCF for DL-ISR reflects the contributions IS scholars can make to learn from complex encodings and architectures. These contributions can be made to the paradigm (e.g., supervised, transfer) or to the strategy (e.g., joint training). With regards to the paradigm, novelties include converting architectures to operate in alternate settings (e.g., supervised to unsupervised), extending existing paradigms to operate on new architectures, or defining new approaches based on computational theories, design theories, SBE theories and/or natural science inspirations from/to the Knowledge Base. Innovations on the learning strategy enhance how an architecture learns from encoded data. Learning strategy novelties include approaches to jointly train multiple model architectures, new loss functions for specialized architectures (tied closely to business objectives), extended objective functions, and others. Each contribution can help optimize, regularize, and effectively execute architecture parameters.

Common issues in this category are threefold. First, engineering procedures (e.g., adjusting batch sizes, learning rates) considered in all DL research can often be mistaken as contributions. Second, variations can often lack proper justifications or rationale. This often occurs when a learning strategy is devised without careful consideration of key domain requirements. Finally, scholars can often loosely couple components (i.e., train them separately) to achieve a larger goal. This specifically occurs when merging conventional ML models with DL strategies.

**DL-ISR Contribution Type 4: System- or Framework Novelty**

The system- or framework-level novelty in the KCF for DL-ISR speaks directly to the long-standing tradition of systems development in IS [39] as well as enable an inclusive research agenda for all IS paradigms. These novelties focus less on each individual component and more on composing multiple techniques into a larger workflow that is greater than the sum of its parts. For example, inputting variables obtained from DL models into econometric models can help further identify causality (a strength of econometric models) within phenomena. DL-based systems can also serve as mechanisms to conduct targeted user evaluations to test for adoption, impact, and governance or study issues of AI trust, bias, or fairness. Two key issues can arise when striving to attain system or framework novelty. First, each component can lack appropriate justification or grounding. This can be mitigated by clearly articulating the strengths, weaknesses, and past examples of success. Second, these works can face significant challenges in evaluation. Carefully designing a workflow requires evaluating each individual component in its entirety.
**DL-ISR Contribution Type 5: Innovation-Accelerating Contributions**

Many researchers re-implementing DL rarely attain the same performances presented in the original paper or are unaware on the best approaches to operate DL for their task (e.g., extracting variables for econometric models). These issues are often due to a lack of details provided in published papers about parameter tuning, datasets, assumptions, etc. The final KCF contribution type provide authors to position their public releases of content or replication studies to help address these issues as contributions. Publicly accessible content such as instruments, code, datasets, and others essential to the core analytics process can allow scholars to quickly build on past work and develop, test, and refine newer contributions (irrespective of IS research paradigm). Knowledge about null results can help scholars learn from past efforts, save time, and avoid fruitless paths [40]. While conducting research replicating DL studies may seem to slow down short-term progress, it is essential for the long-term benefit of an academic discipline for three reasons. First, if a replication does not hold, additional work will aim to identify why [15]. Second, successful replications can point to promising future directions. Finally, executing replications can also help quickly onboard new Ph.D., DBA, MS, and undergraduate students into the field. *AIS Transactions on Replication Research* serves as a vehicle to facilitate scientific replication and theory building in IS.

**Guidelines for Presenting and Positioning DL-ISR**

The KCF provides a novel lens for IS scholars to systematically contribute to DL. However, consistently contributing rigorous and thorough DL research can be accelerated by providing a comprehensive set of guidelines on how to execute DL-ISR. Since DSR is designed to be naturally synergistic with other paradigms of IS research (e.g., behavioral), we propose ten guidelines based in DSR guidelines. These guidelines aim to balance research, practice, and IS tradition to help IS researchers address and navigate some of DL’s unique aspects and challenges to execute research inquiries for DL based on design, behavioral, and/or economic perspectives. The guidelines can be grouped into four categories: Setup, Solution, Study, and Synthesis. We present each category, their guidelines, and their mapping to DSR in Table 5. Since DSR is primarily for technical research, we provide references from top IS journals that have provided guidance on executing scientific research for each guideline to help further ground each guideline.

Setup pertains to the motivation for the proposed DL research. Solution comprises of the DL development and evaluation. Study focuses on reflecting on the setup and solution and positioning the novelty of the study via the KCF. Synthesis focuses on how the proposed DL research can be translated into the application environment it was motivated by. Writing each guideline in a prescriptive manner to help authors and review teams consistently present or evaluate new ideas. The guidelines are written in a manner that does not favor one IS tradition and can therefore help IS scholars different types of contributions across DL-ISR, not just to technical DL design. Consequently, the extent to which the guidelines are followed would likely vary based on the nature and focus of the study (e.g., behavioral studies would have less DL evaluation).
### Table 5. Guidelines for executing Deep Learning for Information Systems Research.

| Category   | Guideline                                                                 | DSR guidelines | References |
|------------|---------------------------------------------------------------------------|----------------|------------|
| Setup      | Motivate and provide a clear problem specification.                        | Problem Relevance | [41, 44, 46] |
|            | Review and summarize key data generating processes and past work to justify deep learning. | Design as a Search Process | [41, 48] |
| Solution   | Systematically present the deep learning processes and how they were formulated and developed. | Design an Artifact | [9, 41] |
|            | Rigorously evaluate the proposed deep learning approach with appropriate technical and non-technical experiments. | Design Evaluation and Research Rigor | [22, 24] |
|            | Provide procedural details of the deep learning approach to facilitate scientific reproducibility. | Research Rigor | [9, 15, 41] |
|            | Articulate clearly the DL contribution via the KCF for DL-ISR.             | Research Contributions | - |
| Study      | Reflect and summarize the key takeaways of the DL process.                | Communication of Research | [50] |
|            | Discuss potential unexpected or unintended consequences of the proposed deep learning process and implementation. |               | [25, 28, 50] |
|            | State any potential ethical concerns in how the deep learning approach was developed and in its potential usage. |               | [25, 28, 51] |
| Synthesis  | Articulate how the proposed deep learning work can be translated into practice or industry. |               | [47, 53] |

**Notes:** DSR, design science research; KCF, knowledge contribution framework; DL-ISR, deep learning-information systems research.

### Setup

**Guideline 1: Motivate and Provide a Clear Problem Specification**

Irrespective of IS tradition, a research project requires a clear problem specification and motivation. For projects with DL, wo levels of details should be included. The first pertains to the higher-level overarching problem motivation (e.g., costs, federal initiatives, and other key statistics) of the key societal, business concern, and/or environment being studied. The second, lower level focuses on the specific problem being studied within the overarching context. Clearly articulating the specific problem being studied is essential for keeping the scope bounded. More importantly, it guides how the overall DL process is formulated (e.g., data, task, and requirements) and evaluated. A common issue is expressing the higher-level overarching problem or the lower level clearly. However, both levels are essential to convincingly convey the scope, scale, importance, and bounds of the study.

**Guideline 2: Review and Summarize Key Data Generating Processes and Past Work to Justify Deep Learning**

Common questions that arise during academic reviews pertain to the data being studied, why DL is needed, the appropriateness and completeness of the data, and features being used. Scholars can proactively address these concerns by reviewing the data-generating processes of the application, past research examining these data, or have strong grounding in SBE related theories [48]. Key details for the data generating processes include summarizing how the data comes into existence, their provenance, data dictionaries, causal linkages, quality, assumptions, uncertainties, and theoretical relationships. Reviewing past work can further summarize how researchers have leveraged these data characteristics for the domain and problem under consideration. Key dimensions of past works that can be examined include motivation, representation, task, architecture, learning strategy, and key results. Well-executed reviews or established SBE theories can ground relevant DL encodings, architectures, learning strategies, paradigms, and evaluations.
Solution

Guideline 3: Systematically Present the Deep Learning Processes and How They Were Formulated and Developed

Contributing knowledge about DL entails summarizing the key design logic of the proposed processes. Key activities include exploring and justifying candidate encodings, architectures, learning strategies based on the data, task, and requirements. Researchers can present the conceptual novelty by presenting key notation, mathematical formulations, pseudocode, end-to-end examples, and side-by-side diagrams that summarize the key differences between the proposed and original approaches. These descriptions should be accompanied by a summary of the technical and non-technical benefits of the proposed approach. Naturally, the depth of these details would vary based on the type of study DL is used for. For example, studies that use DL solely to extract variables for econometric modelling would likely justify their model selection and usage rather than provide a detailed mathematical formulation. Similarly, studies using DL to enhance or complement behavioral methods (e.g., analyzing interview data) can justify model selection and provide detail about how the parameters of the model were varied for their task.

Guideline 4: Rigorously Evaluate the Proposed Deep Learning Approach with Appropriate Technical and Non-Technical Experiments

Guideline 4 necessitates that researchers also demonstrate the validity (i.e., does it solve the problem being posed) of a DL approach through technical or non-technical approaches. Technical evaluations compare the performance (through well-established metrics) of a DL process against prevailing ML and DL benchmarks. Five components comprise a thorough technical evaluation: dataset, model training and testing, model performance benchmarking, post-hoc (i.e., post-model) evaluation, and interpretation and insights (i.e., technical case study). The key aspects of each major component are presented in Table 6.

The scale, scope, breadth, and depth of technical evaluations will vary based on DL contribution type. Scholars who claim their proposed DL approaches have representation or learning level novelties may need to employ more extensive technical demonstrations. Conversely, system- or application-level novelties may rely more on behavioral evaluations to evaluate a DL process. Behavioral evaluations can also help determine whether the proposed technical work solves the higher-level problem in the environment. Possible evaluation approaches include surveys, randomized field experiments, embedded simulations, case studies, interviews, focus groups, questionnaires, and ethnographies. Subjects included in each evaluation type should ideally be drawn from the environment for which the approach is designed for to convincingly illustrate the potential impact (e.g., AI trust, security, and privacy issues) and adoption of the DL approach.

We do not advocate that both technical and non-technical approaches are required to demonstrate the validity, value, and usefulness of the proposed techniques. Non-technical evaluations may not be feasible or may significantly delay the publication of high-impact research. These situations are common for healthcare and cybersecurity applications [53]. Similarly, technical evaluations may not be required if the algorithms being deployed are well-accepted for specific contexts, or if DL is not the centerpiece (e.g., generating variables for an econometric model).
Table 6. Summary of major components in a technical evaluation for DL-based IS research.

| Component | Key aspects | Description | Example(s)** |
|-----------|-------------|-------------|--------------|
| Dataset   | Ground-truth dataset construction | Labelled dataset used for model training and testing | Complete dataset fully labelled by experts |
|           | Train       | Data used to train the algorithm(s) | 80 percent of the ground truth dataset |
|           | Development | Data used to tune the algorithm(s) | 10 percent of the ground truth dataset |
|           | Testing     | Portion of data that is used to test and evaluate algorithm performance | Randomly selected 10 percent of the ground truth dataset |
| Model Training and Testing | Hyperparameter selection | Selecting values to control the learning process | Grid-search, pre-optimized, or trained model |
|           | Training strategy | How the proposed model is trained and the model parameters learned | 10-fold cross validation, hold-out, pre-trained model |
| Model Performance Benchmarking | Performance metric selection | Metrics to evaluate the performance | Accuracy, precision, recall, F1, MAP, MRR |
|           | Evaluation against non-DL models | Proposed DL model vs non-DL-based models | Naïve Bayes, SVM, Decision Tree |
|           | Evaluation against DL models | Proposed DL model vs prevailing DL-based models | CNN, LSTM, RNN, ANN |
| Post-hoc (i.e., post-model) evaluation | Sensitivity or ablation analysis | Internal analysis of DL model to interpret how model components contribute to overall performance | # of layers, activation functions, varying model components, counterfactual analysis |
|           | Convergence speed | How quickly the model converges | Speed, complexity |
|           | Model stability | How stable the model is in training, comparison, etc. | Validation loss, thresholding, statistical significance |
| Interpretation and Insights (Technical case study) | Examples of outperformance | 1-2 instances within the ground-truth dataset that were correctly identified by the proposed method, but missed by the best competing benchmark | Applying approach to relevant datasets within the environment |
|           | Apply proposed DL on unseen data | - | - |

*Note: Post-hoc evaluations are similar to robustness checks in econometrics research. However, running these tests requires significantly higher investment and resources. Therefore, review teams should have a well-grounded request for authors to run these. **Note: ANN, artificial neural network; CNN, convolutional neural network; DL, deep learning; LSTM, long-short term memory; MAP, mean average precision; MRR, mean reciprocal rank; RNN, recurrent neural network; SVM, support vector machine.
Guideline 5: Provide Procedural Details of the Deep Learning Approach to Facilitate Scientific Reproducibility

A key contribution that can be made from studies leveraging DL (technical, behavioral, or economic) is the provision of the code, data, and documentation to verify that the reported performances are true and correct. In cases where such provisions are not possible, authors should ensure their paper provides sufficient details to allow subsequent scholars to replicate the processes and reproduce the same results. This can be done with a combination of pseudocode (for DL-design) and/or a detailed appendix. Key details to disclose include parameters, layers, pre-processing steps, key operations, hardware setups, and programming packages [67].

Guideline 6: Articulate Clearly the DL Contribution via the KCF for DL-ISR

The KCF was intentionally developed to support contributions from multiple IS paradigms. Therefore, clearly summarizing the DL for ISR contributions can manifest itself by positioning the contribution in the KCF. Contributions should be stated in a way that helps the audience understand the work’s core contribution [23] and how future work can build upon the proposed work.

Study

Guideline 7: Reflect and Summarize the Key Takeaways of the Deep Learning Process

Carefully studying the value and key takeaways of the proposed DL processes is essential for realizing the full potential of a DL-ISR study. Studies using existing DL systems to enhance behavioral or economic approaches can summarize the value attained from leveraging DL and practical considerations for future researchers. For studies producing new DL designs, authors should reflect on the generalizable design principles or engineering considerations of their proposed research for other disciplines or domains. These details are important contributions to the knowledge base and can help drive new IS research. Authors should also articulate the boundary conditions of their work to help prevent scope creep or incorrect applications of their method.

Guideline 8: Discuss Potential Unexpected or Unintended Consequences of the Proposed Deep Learning Model and Its Implementation

IS scholars should proactively indicate unexpected or unintended consequences or behaviors (boundary conditions) of their proposed DL processes. Three perspectives can be taken. First, researchers can identify instances where their proposed DL processes behaved in an unexpected fashion. Second, researchers can imagine how their proposed DL process would behave at scale. Finally, researchers can consider executing behavioral studies of new or existing DL artifacts to ascertain AI trust, privacy, and security issues.

Guideline 9: State Any Potential Ethical Concerns in How the Deep Learning Approach Was Developed and In Its Potential Usage

Ethics are increasingly playing a larger role in academia. To this end, IS scholars should carefully consider the potential ethical issues and impacts of DL research. Key questions to consider include how the data were collected, biases implicit in the dataset, post-hoc analysis of the converged DL models to enhance interpretability and how the learning
process behaves, and ethical concerns that might arise when the model is deployed in practice. Each detail can help facilitate the design of just DL systems, governance strategies, incentivize usage, motivate new DL designs, and drive behavioral research on how DL is adopted and used across various conditions (e.g., technological, social, cultural).

Synthesis

Guideline 10: Articulate How the Proposed Deep Learning Work Can Be Translated into Practice or Industry

There is significant focus across the academic landscape on attaining external grant funding from government (e.g., National Science Foundation) or industry. However, attaining funding requires clear articulations of the practical utility and value of a DL-based work (irrespective of IS paradigm). Detailed discussions of how the environment and model outcomes together influence business and societal outcomes can help articulate the translational process.

An Example of a Deep Learning Contribution: Detecting Emerging Threats from the Dark Web

We illustrate a contribution to DL for cybersecurity, specifically for identifying emerging threats from Dark Web for cyber threat intelligence (CTI). We provide background for this research, present DL-based framework, and evaluation and case study results. We map the activities to the schematic and to the ten guidelines and summarize research contributions.

Research Background and Motivation

Cybersecurity practitioners (Application Environment) develop CTI about emerging threats and key threat actors to help protect their critical assets. IS scholars have noted that the "Dark Web" contains significant exploits that can help develop proactive CTI [13, 18, 54]. However, extant approaches for detecting emerging threats have used term frequency, bag-of-words (BoW), or keyword [55] approaches that cannot capture the relationships in a corpus. Word embedding approaches (e.g., word2vec) can address some of these issues. However past studies have applied such techniques synchronically; consequently, they cannot detect emerging threats. These limitations necessitate: (1) an alternate approach to represent hacker forum text; (2) a DL-based approach to generate embeddings from the text representation; (3) and identifying how word embeddings semantically evolve to identify emerging threats.

Proposed Framework and Research Contributions

We design a diachronic graph convolutional autoencoder (D-GCAE; Figure 2) based on text graphs, graph embeddings, and diachronic linguistics (from the knowledge base) to detect emerging threats for CTI.

D-GCAE comprises of three major components:
Figure 2. Proposed D-GCAE for detecting emerging threats. D-GCAE is a contribution to the Knowledge Base.

- **Time-Spell and Exploit Text Graph (ETG) Construction**: Conducting diachronic linguistics requires splitting datasets into equidistant time-spells. For this illustration, we split the time-spells into three months consistent. For each time-spell, we construct a novel ETG (encoding) for exploit text based on a graph-of-words (GoW). Formally, the ETG in each time-spell is denoted as $T = (V,E)$, where $T$ is the overall ETG, $V$ is the set of words in that time-spell, and $E$ is the set of relationships between words if they appear in the same post. Modeling hacker text in this fashion captures a word’s local and global relationships missed by BoW.

- **Graph Convolutional Autoencoder (GCAE)**: Facilitating diachronic analysis requires word embeddings. The prevailing approach for generating embeddings from graphs is the Graph Convolutional Network (GCN) [10]. However, GCN has been used in supervised learning tasks with a ground-truth dataset (often unavailable for hacker forum data), we develop a novel GCAE that incorporates custom GCN operations into an autoencoder [21] to create an embedding for each node at each time-spell in an unsupervised manner (without a priori knowledge). The GCAE is presented in Figure 3.

- The GCAE receives the $N \times N$ adjacency matrix from $T$ and an $N \times N$ identity matrix $I$ to create an embedding for each node by extending the standard autoencoder encoder, $h(x) = \sigma(Wx + b)$, to include graph convolutions as formulated by $h(x) = \sigma(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}HW + b)$, where $\sigma$ is the ReLU activation function, $H$ is the row-wise embedding of the graph nodes in a layer, $W$ is the weight matrix, and $b$ is the bias term. We used the standard adjacency matrix rather than the conventional approach of adding an identity matrix. In this way, if self-loops were to occur, this would not skew the importance of words. The extended encoder creates a low-dimensional embedding of each node. The standard decoder is extended from $\sigma(W^*(h(x)) + b)$ (where $h(x)$ is the encoder’s output embedding) to $\hat{x} = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}})(W^*(\sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}HW + b)) + b)$. Each decoder layer applies these operations to reconstruct the output. The error between the reconstructed output and original input is used for backpropagation.
Diachronic Operations: The final D-GCAE component relies on embedding space alignment and computing semantic displacement. For the former, we align embedding spaces across time-spells while retaining cosine similarity by optimizing

\[ R(t) = \arg\min_{Q} W(t)Q - W(t+1)F, \]

where \( || \cdot ||_F \) denotes the Frobenius norm. Computing semantic displacement (to identify emerging threat terms) relies on calculating the cosine distance of a word’s embedding across time-periods with:

\[ \text{cosine-dist}(w_t, w_{t+\Delta}), \]

where \( w_t \) is the location of a word within the embedding space at time \( t \), and \( w_{t+\Delta} \) is the location of the word after time \( t \).

Results and Discussion: Evaluation and Case Study

We conducted two sets of technical evaluations: (1) GCAE vs word2vec, the prevailing word embedding approaches and (2) GCAE vs prevailing graph embedding algorithms such as LINE, node2vec, DeepWalk, and GraRep [10]. We evaluated embeddings generated from these approaches in a clustering task, as CTI professionals often wish to understand an exploit and its surrounding terms. To conduct both evaluations, we developed a ground-truth dataset with three clusters (30 exploits each): keyloggers, crypters, and remote administration tools. The metrics of homogeneity, completeness, and V-Measure were used to evaluate clustering performance. GCAE embeddings attained a V-Measure of 49.113 percent, significantly higher than word2vec (31.000 percent), LINE (37.756 percent), node2vec (33.635 percent), DeepWalk (26.982 percent), and GraRep (25.180 percent). Interested readers can contact the first author for full details on the datasets and results. We applied the proposed D-GCAE on 235 ransomware exploits between March 11, 2010 and October 20, 2017. We summarize the top 20 words with the highest average semantic shift (new threats) in Table 7.

Each word in the corpus shifted an average amount of 1.2538 in semantic space. Words with the top 20 average shifts relate to specific ransomware functionalities (e.g., “bitcoin” for payment mechanisms, “steal,” “variant” for exploitation). One top word is “infect,” which appears at rank 20 (shift of 1.51857). “Infect” in a total of 19/235 (8.0551 percent)
ransomware posts. When examining how the meaning of “infect” has shifted from 2010 to 2014 to 2017, we noticed that “infect” pertained to system drivers, specifically memory injections to steal driver objects in 2010. In 2014, the term “infect” moves beyond infecting memory processes to damaging executables on a victim’s machine. This enables attackers to alter Windows Explorer settings and monitor user activities. In 2017, “infect” pertained to modifying specific files (e.g., “project.ppt”). This CTI can provide valuable tactical leads (business or societal outcomes) for CTI professionals (application environment), especially in healthcare, an industry widely afflicted with ransomware.

**Summary of Research Contributions and Mapping to DL-ISR Guidelines**

When considering the KCF, the ETG and GCAE are representation level (DL design) contributions as they are new approaches for representing hacker forum text for cybersecurity (ETG) or is a new architecture (GCAE) that learns from a graph structure from the cybersecurity application environment. The overall D-GCAE is a framework contribution, as it integrates multiple components (ETG, GCAE, diachronic operations) into a single workflow to detect emerging threats from Dark Web forums. These contributions are new additions to the knowledge base. With regards to the guidelines the presentation of the D-GCAE is consistent with Guideline 1 with a clear articulation of the problem of identifying emerging threats for proactive CTI, Guideline 2 by summarizing how hacker forums operate and key metadata, Guideline 3 by presenting the DL solution with diagrams and notation, Guideline 4 with a rigorous evaluation of the core DL component of the framework in a manner consistent with best practices in literature, Guideline 5 by providing an option for interested readers to contact the lead author for full model implementation and datasets, and Guideline 6 by positioning the contribution via the KCF. Guidelines 7-10 would likely be further described in a full journal paper. For example, the discussion around the key takeaways of the DL process (Guideline 7) would likely revolve around the generalizable design principles learned from the D-GCAE. Explanations of Guidelines 8 and 9 could revolve around how D-GCAE could be used illicitly (e.g., for surveillance). Finally, discussion Guideline 10 could point to stakeholders (e.g., security operations centers) likely to use D-GCAE and potentially chart a path (e.g., user studies) on how to make it usable.

### Table 7. Top Shifted Words Between 2010-2017 for Ransomware (*average shift per time-spell)*

| Rank | Word      | Amount shifted* | Rank | Word      | Amount shifted* |
|------|-----------|-----------------|------|-----------|-----------------|
| 1    | Initial   | 1.56378         | 11   | Instal    | 1.53015         |
| 2    | Variant   | 1.55563         | 12   | Host      | 1.52990         |
| 3    | Steal     | 1.55430         | 13   | Vari      | 1.52902         |
| 4    | Touch     | 1.55418         | 14   | Strategi  | 1.52778         |
| 5    | Organ     | 1.54652         | 15   | Case      | 1.52593         |
| 6    | Summer    | 1.53727         | 16   | Financi   | 1.52273         |
| 7    | Wolf      | 1.53707         | 17   | August    | 1.52201         |
| 8    | Mine      | 1.53594         | 18   | Major     | 1.51968         |
| 9    | Bitcoin   | 1.53217         | 19   | Establish | 1.51908         |
| 10   | multicompon | 1.53145     | 20   | Infect    | 1.51857         |
Conclusion

Deep learning has rapidly emerged as an essential component to modern AI. However, DL research within IS is still in its nascency. Clear, prescriptive guidance on how IS scholars can approach, examine, and make DL-oriented contributions is essential for ensuring the IS discipline makes timely and high-impact contributions to this rapidly emerging topic. To this end, this paper aimed to make three sets of contributions. First, we systematically summarized the major components of DL in a novel DL-ISR schematic, where technical DL processes are driven by key factors from an Application Environment and/or Knowledge Base. Second, we presented a KCF to provide IS scholars a roadmap on how to clearly position their contributions to DL-ISR. The KCF can help accelerate the rate, depth, and breadth of DL-ISR. Third, we provided a set of guidelines to help IS scholars systematically and scientifically generate rigorous and relevant DL-ISR. Taken together, we believe that these contributions can provide IS scholars an unprecedented ability to advance the scale, scope, and impact of DL research to create a positive societal impact.

Note

1 It is important to delineate between DL research and DL practice. DL research aims to provide new knowledge (i.e., contributions to the knowledge base) about DL or with DL (e.g., to enhance existing behavioral, economic, or technical methods or studies). As such, DL-related research activities could occur within the Standard ML Processes, DL design, and/or business or societal outcomes components. Each concept within these components has the potential to be studied, but the focus of the research activities may differ since each paradigm of IS research has specific details about what research activities should constitute (e.g., iterative science research build, test, evaluate cycles) and how knowledge is abstracted to general insights that are provided to the Knowledge Base. Therefore, we do not provide specific, explicit details about research activities in the schematic and only point to where such activities could occur. DL in practice is largely focused on using and applying existing algorithms to known problems without considering what new knowledge is contributed to the knowledge base (the key difference between DL research and DL practice). Gregor and Hevner [22] refer to standard deployments as “routine designs.”

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This material is based upon work supported by the National Science Foundation under Grant Numbers OAC-1917117 (CICI), CNS-1936370 (SaTC CORE), CNS-1850362 (CRII SaTC), and DGE-2038483 (SaTC-EDU). Y. Chai was also partially supported by the National Natural Science Foundation of China (91846201, 72101079, 71722010) and the Shanghai Data Exchange Cooperative Program (W2021JSZX0052).

Notes on contributors

Sagar Samtani (ssamtani@iu.edu) is an Assistant Professor and Grant Thornton Scholar in the Department of Operations and Decision Technologies at the Kelley School of Business at Indiana
University. He received his Ph.D. from the Artificial Intelligence Lab at the University of Arizona, where he served as a National Science Foundation Scholarship-for-Service Fellow. Dr. Samtani’s research focuses on developing AI-enabled algorithms and systems for cybersecurity and mental health applications. He has published over 60 journal, conference, and workshop papers in venues such as MIS Quarterly, Information Systems Research, Journal of Management Information Systems, ACM Transactions on Privacy and Security, IEEE Transactions on Dependable and Secure Computing, and others. His research has received funding from the NSF and other agencies, and has won several awards for his research. He was inducted into the NSF/CISA CyberCorps SFS Hall of Fame in 2022 and was named by Poets and Quants as a Top 50 Undergraduate Business School Professor the same year. Dr. Samtani’s work has received media attention from outlets such as Miami Herald, Fox, Science Magazine, and AAAS.

**Hongyi Zhu** (hongyi.zhu@utsa.edu) is an Assistant Professor in the Department of Information Systems and Cyber Security at the College of Business at the University of Texas at San Antonio. He received his Ph.D. in Management Information Systems from the University of Arizona. Dr. Zhu has primarily worked on designing advanced mobile analytics for smart home care, focusing on the recognition, extraction, and analysis of subjects’ in-house behaviors from raw mobile sensor data. His work has been published or accepted in journals such as the MIS Quarterly, Journal of Management Information Systems, Journal of Biomedical Informatics, IEEE Intelligent Systems, and others. He has contributed to a variety of projects supported by the National Science Foundation.

**Balaji Padmanabhan** (bp@usf.edu) is the Anderson Professor of Global Management, the Director of the Center for Analytics & Creativity, and a Professor in the School of Information Systems and Decision Sciences at the University of South Florida. He earned his Ph.D. from the Stern School of Business at New York University. Dr. Padmanabhan’s research interests and expertise include AI and machine learning, designing analytics-driven algorithms for business applications, building and evaluating predictive models, patterns discovery in data, business value of analytics, enabling citizen data science and applications of analytics in churn, health care, recommender systems, fraud detection, and elections. His work has been published in the premier computer science and business journals and conferences. He serves as a Senior Editor at MIS Quarterly, Guest Senior Editor at Information Systems Research and Journal on the Association of Information Systems, and an Associate Editor at Management Science, ACM Transactions on MIS, and Big Data.

**Yidong Chai** (chaiyd@hfl.edu.cn; corresponding author) is an Assistant Professor in the School of Management at Heifei University of Technology, China. He received his Ph.D. from the Department of Management Science and Engineering at Tsinghua University, China. Dr. Chai’s research interests include deep learning, text mining, image analysis, and fall risk assessment. His work has been published at MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Information Processing and Management, Knowledge-Based Systems, and other journals and conference proceedings.

**Hsinchun Chen** (hsinchun@arizona.edu) is the University of Arizona Regents’ Professor and Thomas R. Brown Chair professor in Management and Technology. He is also a Fellow of ACM, IEEE, AAAS, and AIS. Dr. Chen served as the lead program director of the Smart and Connected (SCH) Program at the NSF for 2014-2015, a multi-year multi-agency health IT research program of in the United States. He is author/editor of 20 books, 300 SCI journal articles, and 200 refereed conference articles covering digital library, data/text/web mining, business analytics, security informatics and health informatics. Dr. Chen founded the Artificial Intelligence Lab at The University of Arizona in 1989, which has received $600M+ research funding from NSF, NIH, NLM, DOD, DOJ, CIA, DHS, and other agencies (100+ grants, 50+ from NSF). He has served as Editor-in-Chief, Senior Editor or AE of major ACM and IEEE journals and conference/program chair of major conferences. He is also a successful IT entrepreneur. His COPLINK/i2 system for security analytics was commercialized in 2000 and acquired by IBM as its leading government analytics product in 2011. He is a visiting chair professor at several major universities in China (Tsinghua University) and Taiwan (National Taiwan University).
Jay F. Nunamaker Jr. (jnunamaker@cmi.arizona.edu) is Regents and Soldwedel Professor of MIS, Computer Science and Communication, and director of the Center for the Management of Information and the National Center for Border Security and Immigration at the University of Arizona. He received his Ph.D. in Operations Research and Systems Engineering from Case Institute of Technology. Dr. Nunamaker has held a professional engineer’s license since 1965. He was inducted into the Design Science Hall of Fame and received the LEO Award for Lifetime Achievement from the Association for Information Systems. He was featured in the July 1997 issue of Forbes Magazine on technology as one of eight key innovators in information technology. His specialization is in the fields of system analysis and design, collaboration technology, and deception detection. The commercial product GroupSystems ThinkTank, based on his research, is often referred to as the gold standard for structured collaboration systems. He founded the MIS Department at the University of Arizona and served as department head for 18 years.

References

1. Abbasi; A., Sarker; S., and Chiang, R.H.L. Big data research in information systems: Toward an inclusive research agenda. Journal of the Association for Information Systems, 17, 2 (2016), 3.
2. Abukmeil, M.; Ferrari, S.; Genovese, A.; Piuri, V.; and Scotti, F. A survey of unsupervised generative models for exploratory data analysis and representation learning. ACM Computing Surveys, 54, 5 (2021), 1–40.
3. Adamopoulos, P.; Ghose, A.; and Todri, V. The impact of user personality traits on word of mouth: Text-mining social media platforms. Information Systems Research, 29, 3 (2018), 612–640.
4. Aggarwal, C.C. Neural Networks and Deep Learning: A Textbook. Springer, 2018.
5. Ahmad, F.; Abbasi, A.; Kitchens, B.; Adjeroh, D.A.; and Zeng, D. Deep learning for adverse event detection from web search. IEEE Transactions on Knowledge and Data Engineering (2020), 1–1.
6. Anderson, C. The end of theory: The data deluge makes the scientific method obsolete. Wired Magazine, 16, 7 (2008), 16–07.
7. Bengio, Y.; Lecun, Y.; and Hinton, G. Deep Learning for AI. Communications of the ACM, 64, 7 (2021), 58–65.
8. Berente, N.; Gu, B.; Recker, J.; and Santhanam, R. Managing artificial intelligence. MIS Quarterly, 45, 3 (2021), 1433–1450.
9. Burton-Jones, A.; Boh, W.F.; Oborn, E.; and Padmanabhan, B. Editor’s comments: Advancing research transparency at MIS Quarterly: A pluralistic approach. MIS Quarterly, 45, 2 (2021), iii–xviii.
10. Cai, H.; Zheng, V.W.; and Chang, K.C.-C. A comprehensive survey of graph embedding: Problems, techniques, and applications. IEEE Transactions on Knowledge and Data Engineering, 30, 9 (2018), 1616–1637.
11. Calin, O. Deep Learning Architectures: A Mathematical Approach. Springer Nature, 2020.
12. Cao, L. AI in finance: Challenges, Techniques, and Opportunities. ACM Computing Surveys, 55, 3 (2022), 1–38.
13. Chen, H.; Chiang, R.H.L.; and Storey, V.C. Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36, 4 (2012), 1165–1188.
14. De-Arteaga, M.; Herlands, W.; Neill, D.B.; and Dubrawski, A. Machine learning for the developing world. ACM Transactions on Management Information Systems, 9, 2 (2018), 1–14.
15. Dennis, A.; Brown, S.A.; Wells, T.M.; and Rai, A. Editor’s comments: Replication crisis or replication reassurance: Results of the IS replication project. MIS Quarterly, 44, 3 (2020), iii–x.
16. Deodhar, M.; Ghosh, J.; Saar-Tsehansky, M.; and Keshari, V. Active learning with multiple localized regression models. INFORMS Journal on Computing, 29, 3 (2017), 503–522.
17. Du, M.; Liu, N.; and Hu, X. Techniques for interpretable machine learning. Communications of the ACM, 63, 1 (2019), 68–77.
18. Ebrahimi, M.; Nunamaker, J.F.; and Chen, H. Semi-supervised cyber threat identification in dark net markets: A transductive and deep learning approach. Journal of Management Information Systems, 37, 3 (July 2020), 694–722.
19. Foster, D. Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play. O'Reilly Media, Inc., 2019.
20. Glassner, A. Deep Learning: A Visual Approach. No Starch Press, 2021.
21. Goodfellow, I.; Bengio, Y.; and Courville, A. Deep Learning. MIT Press, 2016.
22. Gregor, S.; and Hevner, A.R. Positioning and presenting design science research for maximum impact. MIS Quarterly, 37, 2 (2013), 337–355.
23. Gupta, A. Editorial—Traits of successful research contributions for publication in ISR: Some thoughts for authors and reviewers. Information Systems Research, 29, 4 (2018), 779–786.
24. Hevner, A.R.; March, S.T.; Park, J.; and Ram, S. Design science in information systems research. MIS Quarterly, 28, 1 (2004), 75–105.
25. Jain, H.; Padmanabhan, B.; Pavlou, P.A.; and Raghu, T.S. Editorial for the special section on humans, algorithms, and augmented intelligence: The future of work, organizations, and society. Information Systems Research, 32, 3 (2021), 675–687.
26. Klein, H.K.; and Myers, M.D. A set of principles for conducting and evaluating interpretive field studies in information systems. MIS Quarterly, 23, 1 (1999), 67–93.
27. Kokkodis, M. Dynamic, multidimensional, and skillset-specific reputation systems for online work. Information Systems Research, 32, 3 (2021), 688–712.
28. Kordzadeh, N.; and Ghasemaghaei, M. Algorithmic bias: Review, synthesis, and future research directions. European Journal of Information Systems, (2021), 1–22.
29. Krohn, J.; Beyleved, G.; and Bassens, A. Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence. Addison-Wesley Professional, 2020.
30. LeCun, Y.; Bengio, Y.; and Hinton, G. Deep learning. Nature, 521, 7553 (May 2015), 436–444.
31. Lee, A.S. and Baskerville, R.L. Generalizing generalizability in information systems research. Information Systems Research, 14, 3 (2003), 221–243.
32. Li, J.; Larsen, K.; and Abbasi, A. Theory on: a design framework and system for unlocking behavioral knowledge through ontology learning. MIS Quarterly, 44, 4 (2020), 1733–1772.
33. Liu, X.; Zhang, B.; Susarla, A.; and Padman, R. Go to YouTube and call me in the morning: Use of social media for chronic conditions. MIS Quarterly, 44, 1 (2020), 257–283.
34. Long, M.; Cao, Y.; Wang, J.; and Jordan, M. Learning transferable features with deep adaptation networks. In F. Bach and D. Blei (eds.), Proceedings of the 32nd International Conference on Machine Learning, PMLR, Lille, France, 2015, pp. 97–105.
35. Lyttinen, K.; Nickerson, J.V.; and King, J.L. Metahuman systems = humans + machines that learn. Journal of Information Technology Impact, 36, 4 (2021), 427–445.
36. Martinez-Plumed, F.; Contreras-Ochando, L.; Ferri, C.; et al. CRISP-DM twenty years later: From data mining processes to data science trajectories. IEEE Transactions on Knowledge and Data Engineering, 33, 8 (2021), 3048–3061.
37. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; and Terzopoulos, D. Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, (2021), 1–20.
38. Mnih, V.; Kavukcuoglu, K.; Silver, D.; et al. Human-level control through deep reinforcement learning. Nature, 518, 7540 (2015), 529–533.
39. Nunamaker, J.; Chen, M.; and Purdin, T. Systems development in information systems research. Journal of Management Information Systems 7, 1991, 89–106.
40. Nunamaker, J.F.; University of Arizona, Twyman, N.W.; et al. Creating high-value real-world impact through systematic programs of research. MIS Quarterly, 41, 2 (2017), 335–51.
41. Padmanabhan, B.; Fang, X.; Sahoo, N.; and Burton-Jones, A. Machine learning in information systems research. MIS Quarterly, 46, 1 (2022), iii–xix.
42. Raff, E. Inside Deep Learning: Math, Algorithms, Models. Manning, 2022.
43. Rai, A. Editor’s comments: Diversity of design science research. MIS Quarterly, 41, 1 (2017), iii–xviii.
44. Rai, A. Seeing the forest for the trees. MIS Quarterly, 41, 4 (2017), iii–vii.
45. Rai, A. Editor’s comments: Beyond outdated labels: The blending of IS research traditions. MIS Quarterly, 42, 1 (2018), iii–vi.
46. Rai, A. The first few pages. MIS Quarterly, 42, 2 (2018), iii–vi.
47. Rai, A. Engaged scholarship: Research with practice for impact. *MIS Quarterly*, 43, 2 (2019), iii–viii.
48. Rai, A. Editor’s comments: Proactively attending to uncertainty in IS research. *Management Information Systems Quarterly*, 44, 1 (2020), iii–vii.
49. Rai, A. Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48, 1 (January 2020), 137–141.
50. Rai, A. Proactively attending to uncertainty in IS research. *MIS Quarterly*, 44, 1 (March 2020), iii–vii.
51. Rai, A., Constantinides, P., and Sarker, S. Next generation digital platforms:: Toward human-AI hybrids. *MIS Quarterly*, 43, 1 (2019), iii–ix.
52. Ramsundar, B.; Eastman, P.; Walters, P.; and Pande, V. *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More*. O’Reilly Media, Inc., 2019.
53. Saar-Tsechansky, M. The business of business data science in IS journals. *MIS Quarterly*, 39, 4 (2015), iiivi.
54. Samtani, S.; Kantarcioglu, M.; and Chen, H. Trailblazing the artificial intelligence for cybersecurity discipline: A multi-disciplinary research roadmap. *ACM Transactions on Management Information Systems*, 11, 4 (2020), 1–19.
55. Samtani, S.; Zhu, H.; and Chen, H. Proactively identifying emerging hacker threats from the dark web: A diachronic graph embedding framework (D-GEF). *ACM Transactions on Privacy and Security*, 23, 4 (2020), 1–33.
56. Schuetzler, R.M.; Grimes, G.M.; and Scott Giboney, J. The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, 37, 3 (2020), 875–900.
57. Shmueli, G.; and Koppius, O.R. Predictive analytics in information systems research. *MIS Quarterly*, 35, 3 (2011), 553–572.
58. Tomašev, N.; Glorot, X.; Rae, J.W.; et al. A Clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*, 572, 7767 (2019), 116–119.
59. Venkatesh, V.; Brown, S.A.; and Bala, H. Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, 37, 1 (2013), 21–54.
60. Wiesche, M.; Münch, T.U.; Jurisch, M.C.; et al. Grounded theory methodology in information systems research. *MIS Quarterly*, 41, 3 (2017), 685–701.
61. Xu, Y.; Pan, S.J.; Xiong, H.; et al. A unified framework for metric transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 29, 6 (2017), 1158–1171.
62. Yang, Q.; Liu, Y.; Chen, T.; and Tong, Y. Federated machine learning: concept and applications. *ACM Transactions on Intelligent Systems Technology*, 10, 2 (2019), 1–19.
63. Zhang, C.; ButePAGE, J.; KJellstrom, H.; and Mandt, S. Advances in variational inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41, 8 (2019), 2008-2026.
64. Zhang, S.; Yao, L.; Sun, A.; and Tay, Y. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52, 1 (2019), 1–38.
65. Zhang, W.; and Ram, S. A comprehensive analysis of triggers and risk factors for asthma based on machine learning and large heterogeneous data sources. *MIS Quarterly*, 44, 1 (2020), 305–349.
66. Zhang, Y.; and Yang, Q. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, (2021), 1-1.
67. Zhu, H.; Samtani, S.; Brown, R.; and Chen, H. A deep learning approach for recognizing activity of daily living (ADL) for senior care: Exploiting interaction dependency and temporal patterns. *MIS Quarterly*, 45, 2 (June 2021), 859–896.
68. Zhu, H.; Samtani, S.; Chen, H.; and Nunamaker, J.F. Human identification for activities of daily living: A deep transfer learning approach. *Journal of Management Information Systems*, 37, 2 (April 2020), 457–483.
69. Chen, K.; Zhang, D.; Yao, L.; Guo, B.; Yu, Z.; and Liu, Y. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, 54, 4 (2021), 1–40.