EDITOR'S COMMENTS

Machine Learning in Information Systems Research

By: Balaji Padmanabhan, Senior Editor
Xiao Fang, Associate Editor Emeritus
Nachiketa Sahoo, Associate Editor
Andrew Burton-Jones, Editor-in-Chief

A key question in the mind of every author submitting a paper to a journal is: What are editors looking for when they assess a paper like mine? During 2022, we will use the Editor's Comments to provide answers to this question for several genres of research that are submitted regularly to *MISQ* and where we believe that authors could benefit from more direction. We hope our answers will help authors to see how our editors think about these genres, how they assess work, what excites them, and what distresses them.

For each editorial, we are following a three-staged process. The editor-in-chief (Andrew) first identifies a select group of editors with substantial expertise with the genre/method to run a masterclass on the topic for our editorial board. This group of editors is chosen to include scholars with different amounts of experience on the board to allow for differences in opinion and perspective.[1] Based on the learning and feedback from that session, the same group then runs an online seminar for authors. We then incorporate the learning and feedback from that session, along with the prior one, to write the editorial.

Having a healthy rotation of editors over time is an important feature of any top journal. As a result, what you will gain from these editorials is advice from a very experienced group of editors. The editorials cannot offer and are not intended to offer the "one true view" of the topic. If you read this editorial and find that your view of the topic differs, please do not worry. In such cases, it can still be useful to know how leading editors view the same topic, even if their perspectives differ from yours.

Having given this context, we are now excited to get into the first genre of the series: machine learning (ML) research. ML plays an increasingly important role in IS research. We can broadly distinguish between two types of ML research in our field (which can also be combined). One type seeks to *study* ML-related phenomena in a particular context, such as how ML is developed, used, and to what effect in a particular organization or industry. Such research can use a range of methods, such as surveys, simulations, and ethnography, among others. Our special issue on *Managing AI* exemplifies this approach (Berente et al. 2021). The second type of ML research in IS seeks to *apply and/or extend ML* itself to make contributions, whether by contributing new ML methods, improving our understanding of IS phenomena through analyses enabled by ML, or advancing our knowledge of complex systems through understanding the role of ML in those systems. It is this second category that we will be covering in this editorial.

ML in Information Systems Research

The last two decades have been a golden era for machine learning (ML), leading to a plethora of exciting applications across many domains, including marketing (Domingos et al. 2001), health (Tomasev et al. 2019), social media (Chen et al. 2009), science (Butler et al. 2018), politics (Padmanabhan and Barfar 2021), and even art (Yeh et al. 2017). Not surprisingly, IS researchers have been particularly active, exploring innovative ideas for ML within mainstream information systems contexts and in broader applications in business and society (Sahoo et al. 2012; Meyer et al. 2014; Shin et al. 2019; Liebman et al. 2019; Fang and Hu 2016; Gorgoglione et al. 2016; Malgonde et al. 2020).

The interest in ML, and more broadly in artificial intelligence (AI), is only likely to increase over the next few years as these technologies continue to move from "big tech" into most organizations. As we see this play out in practice, we will continue to see a parallel phenomenon play out in the IS research community, which is the pursuit of new ideas related to how ML will impact organizations. Judging by recent history, this is likely to generate a significant volume of new ideas produced by our

[1] Balaji Padmanabhan is an SE with two prior AE terms, Xiao Fang recently rotated off the board after three AE terms (and received *MISQ*'s Outstanding AE Award in 2021), and Nachiketa Sahoo is serving his first AE term.

community. Over the years, IS research has seen specific inflection points, often coinciding with high-impact new technologies that are transformational. The WWW and e-commerce were especially influential. Machine learning and Artificial Intelligence can be just as significant.

Yet, the road ahead is not without uncertainties and challenges with respect to how IS researchers can most effectively contribute. Can IS researchers develop new methods here (or is that best left to computer scientists)? If our researchers focus on applications of ML and AI, will the work be perceived as "too applied" to make any generalizable contribution? Is some of this research even "IS" at all—how important is this to ask, and how should this guide authors in their own work? Are we best off modeling these as black boxes and just studying the phenomena around these technologies?[2]

Given this context, this editorial offers guidance as to how IS researchers can exploit some of the rich opportunities ahead. Guided by our editorial experience with handling ML-related submissions, we discuss three types of ML-related contributions that IS researchers have made. Our discussion here is not meant to be exhaustive in terms of identifying all possible contributions IS researchers *can* make. Rather, it is reflective of the types of contributions we have handled as editors over the last decade. We follow this with a broader discussion offering some guidance to authors as they continue to develop their ML-related ideas into papers that can make impactful contributions.

Some Background

Humans learn through experience; this is often by examples or simply interacting with the environment. Inspired by this, the field of machine learning has been captivated by the question: Can computers learn (Mitchell 1997)? This question has launched decades of research that have resulted in numerous algorithms that can "learn." From this perspective, "a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" (Mitchell 1997, p. 2).

Over the last few decades, this experience E, was most often measured based on exposure to "training data" (e.g., large databases of consumer records, images, etc.). Two common classes of techniques here fall under the categorizations of "supervised learning" or "unsupervised learning," based on whether the training data that was provided to the machine had observed labels (or target variables)—in which case the algorithms are "supervised"—or whether the data was just presented as examples without a target variable (e.g., a database of news articles without any class associated with each news article).

There are many variations of these basic learning paradigms based on whether the data has some labels (*semi-supervised learning*, Zhu and Goldberg 2009); whether labels—or just more training data—can be acquired at a cost (*active learning*, Cohn et al. 1996); whether the goal is to learn multiple models simultaneously when there is some underlying synergy of doing so (*multi-task classification*, Caruana 1997); whether the goal is to learn from raw data in its organic form using deep connectionist architectures (*deep learning*, LeCun et al. 2015); whether there are opportunities to leverage vast amounts of unlabeled data by creating supervised learning tasks from them (*self-supervised learning*, Liu et al. 2021); whether the goal is to exploit a model learned from one domain to train a new model in a related, but different, domain (*transfer learning*, Torrey and Shavlik 2010); or whether the data needs to be processed locally for computational or privacy concerns (*federated learning*, Li et al. 2020). These learning paradigms are not exhaustive but offered as examples of the kinds of rich variations in the machine learning literature over the years.

More recently, there has been renewed interest in *reinforcement learning* (Sutton and Barto 2018) algorithms that also learn from experience, but where the experience is based on interactions with an external environment that provides "rewards" based on actions taken by a learning agent. A turning point for this paradigm was the work by the DeepMind team that created AlphaGo, a program that combined deep learning with reinforcement learning to learn how to play a game with minimal instructions but that did so with such expertise that it beat the world champions at the game after just a few weeks of training.

Historically, the computer science community produced the core ML research. However, as ML has permeated disciplines such as business or health, researchers from those communities are increasingly advancing the state of the art in ML. Such research tends to be inspired by important problems in those specific areas, and often results in contributions that are at the intersections of

[2] Having worked in this area collectively for almost five decades, our answer to these questions is: (1) Yes, we can; (2) Not necessarily; (3) Very important; (4) No.

the applying discipline and core ML. Given their technology backgrounds, IS researchers have been particularly active here, bringing to bear their expertise in systems and business to their investigations about how ML can be applied in various business and social domains. In the process of doing so, IS researchers have often chosen to publish in either core ML venues or in the mainstream IS journals. Often the decision of where to submit such research has been guided by multiple considerations, including: (1) academic considerations, i.e., where will the research have the best impact? (2) cultural/structural considerations, i.e., which journal(s) have the necessary review teams, and/or which ones are open enough to embrace work that differs from "traditional" IS papers; and (3) pragmatic considerations, i.e., whether the research is likely to "count" in a business school, particularly for promotion and tenure considerations.

Table 1 presents some early examples of machine learning research published in mainstream IS journals.[3] In most cases, such papers have had specific IS/business contexts that differentiated these ideas from those that tend to get published in core ML venues. We will likely continue to see more such papers in our journals and review teams are most certainly looking for discipline-specific hooks in submissions to discipline-specific journals.

Table 1. Some Early Examples of Applications-Inspired ML Research in Leading IS Journals

Reference	Application & ML paradigm	Journal
Purao et al. (2003)	Bringing ML ideas into conceptual design in information systems development	*Information Systems Research*
Padmanabhan & Tuzhilin (2003)	Ecommerce-inspired ideas for integrating optimization and machine learning ideas to develop hybrid systems	*Management Science*
Saar-Tsechansky & Provost (2007)	Marketing-inspired development of decision-centric active learning ideas	*Information Systems Research*
Abbasi & Chen (2008)	Supervised and unsupervised learning for text analysis in computer-mediated communication	*MIS Quarterly*
Adomavicius et al. (2011)	Use-inspired query language for customizing recommender systems	*Information Systems Research*

Still, given their close ties to the ML community, many IS researchers pursuing ML-related research will likely continue to be active at leading computer science and ML conferences. This leads to a natural cross-pollination of ideas. There are many new ideas in ML that are currently making their way into our field through this organic process. Some notable ones that are currently appearing in submissions to IS journals include ideas such as causal ML (Pearl 2019); deep reinforcement learning (Henderson et al. 2018); attention mechanisms (Vaswani et al. 2017); mechanisms for fairness, explainability, and transparency (Caruana et al. 2020); learning from data organically represented as graphs (Wu et al. 2020); large-scale language models (Floridi and Chiriatti 2020); constructing and using knowledge graphs (Noy et al, 2019); augmented intelligence (Jain et al. 2021); auto-ML (LeDell and Poirier 2020); federated learning (Li et al. 2020); geometric deep learning (Bronstein et al. 2017); and transfer learning (Torrey and Shavlik 2010). It is likely that some, more than others, in this list will see greater traction due to their fit with the current discourses in the IS literature. In addition to bringing new ideas from ML into IS discourse, IS researchers have made unique contributions to the core ML literature too, e.g., decision-centric active learning (Saar-Tsechansky & Provost 2007) and recommender systems (Adomavicius and Tuzhilin 2005).

ML in IS Research: Types of Contributions

What kinds of IS contributions can researchers in our community make in the area of ML? This is a challenging question, for no single categorization is likely to do justice to the depth and range of work in our field. That said, as editors who have handled many ML-related submissions to our journals over the last decade, we have broadly seen three types of contributions that researchers have successfully made. Below we discuss these three types—not for the sake of being comprehensive or to draw rigid boundaries, but simply to help researchers to learn from our experience. We stress that these types are not exhaustive; neither are

[3] Liang (1988) was actually the first to mention ML in *MISQ* in his study of group decision support systems. While we do not include pre-2000 studies in Table 1 because their contexts are further afield from today's settings, the fact that Liang was the first to mention ML in *MISQ* is a fitting reminder of how far ahead of the field he often was (https://aisnet.org/news/566283/AIS-Mourns-the-Loss-of-Past-President-T.P.-Liang.htm).

they mutually exclusive since there may be contributions that span types. We merely believe they can offer a good starting point for consideration when researchers submit their ML work to *MISQ*. The three types are (1) ML methods development, (2) understanding phenomena using ML, and (3) ML within complex systems. We explain each of these below, and in doing so, offer suggestions for authors to consider.

Type I: ML Methods Development

As a sociotechnical discipline, the IS field focuses on the interaction between technology and its business and social contexts (Lee 1999; Niederman and March 2012). ML is a technology that is transforming business and society. Therefore, one contribution that IS researchers are well-positioned to make is to design ML models and algorithms[4] to solve business and societal problems. We refer to such ML research in IS as Type I ML research.

Type I ML research belongs to the computational design science genre, which is "concerned with solving business and societal problems by developing computational models and algorithms" (Rai 2017a, p. iii). Along this line of research, IS researchers have developed ML models and algorithms that solve diverse business and societal problems, in healthcare, finance, social networks, cybersecurity, privacy, and misinformation (Menon and Sarkar 2016; Li et al. 2017; Hendershott et al. 2017; Fang et al. 2021; Samtani et al., in press).

Motivated by unique challenges at the intersection of ML and its business and social environments, Type I ML research aims to make *methodological contributions*. To explain what we mean by "methodological contributions," we refer to the design science paradigm because this research falls within it. The contribution of design science research is an IT artifact in the form of construct, model, method (algorithm), or instantiation (implemented system) (Hevner et al. 2004). Most often, designed IT artifacts are models and algorithms as they are "rarely full-grown information systems" (Hevner et al. 2004, p. 83). Therefore, the methodological contributions of Type I ML research tend to be novel ML models and algorithms developed to solve important business and societal problems.

Methodological contributions are valued strongly by *MISQ* and other journals (Hevner et al. 2004; Rai 2017a; Gupta 2018; Simchi-Levi 2020).[5] For example, *Information Systems Research* values papers that make "methodological, computational, and design contributions" (Gupta 2018, p. 781) and the IS department at *Management Science* values "the development of predictive analytics that clearly combine a methodological advance with an important and novel managerial application" (Simchi-Levi 2020, p. 1).

Figure 1 depicts Type I ML research in relation to the business and social environments as well as the ML technology. The relevance of Type I ML research lies in its objective of solving important problems arising from the business and social environments and the contribution of Type I ML research consists of its developed ML models and algorithms that add to the knowledge base of ML technology (Hevner et al. 2004).

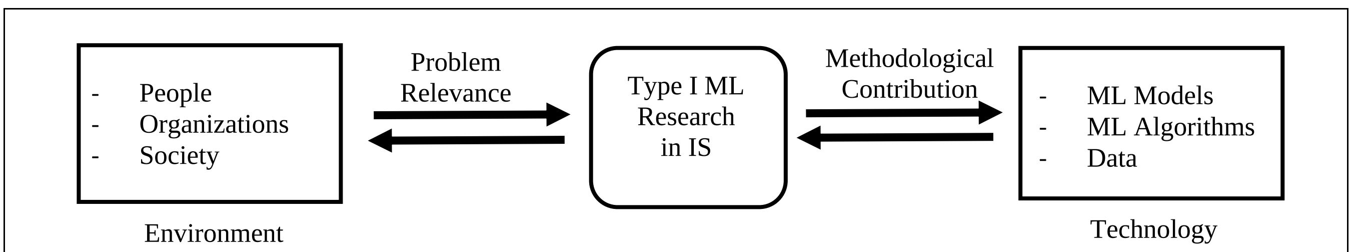

Figure 1. Type I Machine Learning Research in Information Systems

[4] In ML a "model" is what an "algorithm" learns from data. For example, a specific deep learning architecture provides a representation of a model. An algorithm is a computational procedure that learns the best such model from data.

[5] At *MISQ*, we offer three manuscript categories that are relevant for methodological contributions. The Methods Article category is relevant for research where the focal method is a method used by *researchers*. In contrast, the Research Article and Research Note categories are relevant where the focal method is a method used in *practice*. The distinction between the Research Article and Research Note depends on the amount of work required to define the problem-space and contribution. If the research problem is well-known and the intended contribution is clear and can be concisely demonstrated, then the work is more likely to be suitable for the concise format of a Research Note.

Our discussion of methodological contributions in Type I ML research starts from the "Technology" box in Figure 1. As shown, the knowledge base of ML technology is composed of existing ML models and algorithms, as well as those developed for processing data and improving data quality. We list data in parallel to ML models and algorithms because data is the foundation of ML models and algorithms, and the quality of data influences their effectiveness. Data quality has been a central IS research topic for decades (Wang and Strong 1996; Lukyanenko et al. 2019) and is key to building resilient data infrastructure for effective and fair use of data (Stoyanovich et al. 2020; Sadiq et al. 2022). As a result, there are many opportunities for IS researchers to develop ML models and algorithms to address data quality. For instance, Fang et al. (2013) target the timeliness of data and develop a model to decide when to re-run an ML algorithm, and Xu et al. (2021) propose a deep learning method to impute missing values in crowdsourced data to improve data completeness.

To achieve methodological contributions, authors must demonstrate the novelty of their proposed ML model or algorithm compared to existing ones in the knowledge base. This requires authors to provide a sufficient review of existing relevant ML models and algorithms (the left-facing arrow from the "Technology" box in Figure 1). Relevant ML models and algorithms include those *developed for* the problem at hand as well as general-purpose ML models and algorithms that can be *adapted* to solve the problem at hand. The former usually appear in business and social science journals (e.g., IS journals) while the latter can be found in ML outlets in computer science (CS) (e.g., top ML conferences).

Given the fast growth of the ML literature, authors (and reviewers, editors) must keep up with the latest developments. In particular, authors must show the methodological novelty of their developed ML model or algorithm in comparison to reviewed ML models and algorithms, thereby making methodological contributions to the knowledge base (the right-facing arrow to the "Technology" box in Figure 1). Just *applying* existing ML models and algorithms to business and societal problems (namely routine design) usually does not make methodological contributions (Hevner et al. 2004; Gregor and Hevner 2013). In addition, methodological novelties are designed to achieve performance gains. Thus, it is important to show the performance advantage of a proposed ML model or algorithm over well-justified benchmarks. This could be achieved by analytically proving its advantage. More often, it is done empirically by showing its performance advantage over benchmarks with real or simulated data.[6,7] For example, ablation studies are particularly useful, as they tease out the contribution of each novel component of a proposed ML model or algorithm to its performance advantage over benchmarks (see He et al., 2019 for an example ablation study).

Design science research builds "IT artifact[s] in context" (Niederman and March 2012, p. 2). Accordingly, Type I ML research develops novel ML models and algorithms for important problems arising from the business and social environments (the right-facing arrow from the "Environment" box in Figure 1). Hevner et al. (2004) define a problem as the differences between a goal state and the current state of a business or social application/system. These differences are often discovered from theories and findings in business and social sciences. They constitute conceptual novelties (advances) of a Type I ML study, which in turn guide the design of a novel ML model or algorithm and the realization of methodological novelties. For example, when Fang and Hu (2018) developed an ML method to predict top persuaders in a social network, the current state of top persuader prediction predominantly focused on social influence. However, eminent social network theories suggest three forces central to social persuasion, including social influence, entity similarity, and structural equivalence. Therefore, the conceptual novelty of their study is the consideration of all three forces for top persuader prediction, which informs their design of a novel ML method to realize this conceptual novelty.

Type I ML research contributes to the knowledge base of business and social sciences with a novel ML model or algorithm that can solve an important business or societal problem more effectively (the left-facing arrow to the "Environment" box in Figure 1). If deployed in practice, the proposed ML model or algorithm could produce significant business value by reducing cost, increasing profit, or enabling new business models. Therefore, it can be desirable to evaluate the developed ML model or algorithm through a case study, which can demonstrate its superior value over benchmarks in an important business setting.

Type I ML research offers many opportunities for IS research. Although ML has been employed in other business fields, these fields mostly use *existing* ML models and algorithms rather than engaging in Type I ML research. Type I ML research also differs from ML research in the CS field because the confluence of ML and its business and social context is central, whereas technology

[6] Type I ML research could be evaluated using a real-world dataset, or a carefully designed simulation, or a combination of both. The evaluation choice depends on which one can most convincingly demonstrate the novelty and validity of the proposed ML model or algorithm.
[7] While it is important to empirically compare a proposed ML model or algorithm with benchmarks, some other empirical research evaluation paradigms may be less relevant. For example, concerns of endogenous variables are less important for predictive models, unless attaching causal interpretations to corresponding model parameters.

(not context) is central in CS-oriented research. This stronger connection to domain-specific work in business and social sciences is reflected in a thorough review of relevant domain-specific literature and a deep understanding of the business/societal problems to be solved. It also shows up in how ML methods are developed and evaluated. Type I ML research designs context-specific ML methods that incorporate unique characteristics of its business and social domains and often evaluates them using both common ML metrics (e.g., AUC) and domain-specific metrics (e.g., profit).[8] IS researchers are well-positioned to conduct Type I ML research and we encourage them to set their sights high: to produce high impact research that helps to solve critical business and societal problems and tackles the world's grand challenges (Ram and Goes 2021).

Type II: Understanding Phenomena using ML

Studies in this category do not seek to contribute a new ML method but seek to understand phenomena using ML. Three subtypes of studies have been popular: measuring causal effects, proposing domain-specific statistical models, and structural econometric modeling. For each one, we describe opportunities, offer guidance, and highlight pitfalls.

ML for Causal Inference

ML can help measure the effect of a treatment on an outcome in new and challenging domains. One common pattern is to predict a variable of interest (e.g., with information from rich and abundant unstructured data) and then use it as an explanatory variable in an econometric model. For example, Archak et al. (2011) used natural language processing (NLP) to estimate feature assessments in product reviews and then measured the effect of these assessments on subsequent product demand at an online retailer. Zhang et al. (in press) used image classifiers to analyze pictures from Airbnb to detect image quality and room type and then measured the effect of these features on property occupancy.

While such approaches have opened exciting research opportunities, it is important to be aware of some risks and best practices. Measurement errors in explanatory variables in an econometric model can bias the estimates for all the variables, not just the ones with errors. Therefore, it is important to validate the predicted variables in the specific dataset that is being used for the study and report the results. It could require collecting labels from human experts on data subsets. The process should be carefully described.

Fortunately, training and evaluation of predictive tools provide an estimate of the error. This can allow researchers to (1) correct such bias for modest errors, and (2) detect cases in which the predictions are too imprecise and thus draw more qualified conclusions (Yang et al. 2018). Of course, beyond such validation and correction, traditional empirical research standards for causal inference still apply, e.g., regarding theory, identification strategy, and exogeneity assumptions (Angrist and Pischke 2008).

ML is also used to overcome various statistical challenges in measuring causal effects. For example, when the number of covariates is large and of the same order as the number of observations, it is difficult to estimate standard linear regression and carry out inference (Johnstone and Titterington 2009). However, ML-based predictive models can be used to remove the effect of control variables (confounders) from the outcome and treatment variables (Belloni et al. 2013; Chernozhukov et al. 2018). Thereafter, a flexible ML model such as random forest can even learn the heterogeneous effects of treatment at various values of covariates, which can allow personalized interventions (Oprescu et al. 2019; Wager and Athey 2018). Predictive tools can also help estimate treatment effects using instrumental variables for high dimensional datasets with unknown functional dependencies between outcomes and covariates when there are unobserved confounders (Hartford et al. 2017; Newey and Powell 2003). The use of ML for causal inference is an active and evolving area of research. It is helpful to stay open to newer methods and apply them to generate new insights in settings that were previously not accessible to traditional econometric modeling.

---

[8] We are aware of recent ML method-centric work in marketing and operations. In addition, given the close interaction between the ML community in the CS field and tech firms such as Google, Microsoft, and Meta, we have also witnessed examples of ML work motivated by specific problems in these firms. Given these recent trends, we do see an overlap between these kinds of work in other fields and Type I ML Research in the IS field. However, Type I ML research in IS still differs from most work in these other fields such as marketing and CS because of its focus on the interaction between the business and social domains and the ML technology. Such identity of Type I ML research is inherited and specialized from the sociotechnical identity of the IS discipline (Lee 1999; Niederman and March 2012).

However, here are some pitfalls to avoid, particularly for research using predicted variables in econometric models:

1. **Accepting predictions because the model has been validated elsewhere.** This is risky because (1) the performance of a predictive model can vary across datasets, and (2) even a small amount of error in an explanatory variable (because the model used to predict it was not validated on this dataset) can bias conclusions.

2. **Ignoring cleaner alternatives.** Is there a cleaner observed variable that captures the underlying construct? For example, if a study seeks to measure the effect of helpful content in product reviews, we may first think of using NLP to derive a measure of helpfulness. However, there are review platforms that collect helpfulness votes directly from consumers who read and use such reviews to make product purchase decisions. Such data can directly provide the required variables.

3. **Claiming ambiguous contributions**. Sometimes studies appear to claim to contribute a predictive model *and* a causal effect measurement. The researchers may have had to solve a nontrivial prediction problem as well as a causal inference problem, hence an urge to claim both. In our experience, it is better to have only one clear primary contribution because these two types of contribution have different implications for the conduct (design and analysis), presentation, and evaluation of the research (Shmueli 2010). The main contribution of papers using predicted variables in econometric models is measuring causal effects.

Domain-Specific Custom Statistical Models

Researchers can use ML to construct and estimate innovative statistical models based on their domain knowledge and understanding of a context to accurately describe a phenomenon. Often such works adapt and extend existing ML models and use existing techniques to estimate them. While this subtype is similar to some Type I ML Research, the focus here is to contribute substantive insight into a domain, rather than to contribute a method for general use. For example, Xu et al. (2014) propose a mutually exciting point process to model the influences of multiple sources of advertisement exposures to precisely attribute purchases to different types of online ads. They estimate their model using the MCMC approach—a popular estimation technique in statistical ML.

Here are a few suggestions to make an impactful contribution in this subtype of work:

1. Explain why existing statistical models or their simple variations cannot accurately model the phenomenon of interest and the consequence of applying them.

2. Justify the modeling choices ex ante. Discuss considered alternatives as well as specific extensions from the state of the art (instead of presenting the entire model as the contribution) to help readers better understand and contextualize the work.

3. Compare the proposed model with state-of-the-art alternatives (using AIC/BIC, out-of-sample prediction, and other appropriate metrics). The quality of the alternatives is more important than the quantity: a small set of well-regarded recent benchmark models potentially taking different approaches that fit the setting could suffice.

4. Highlight the new insight resulting from the proposed model, since the primary objective of this subtype of study is to better understand a domain.

5. Avoid overemphasizing estimation—it is not the most important contribution for this subtype of papers. Briefly discuss unique estimation issues, use state-of-the-art tools when possible, and provide the details necessary for replication (hyperparameter settings, search procedure, etc.) in supplemental/transparency material.

Structural Econometric Models

The third subtype of work is structural econometric models—statistical models based on optimizing behaviors of agents. One strength of this approach is they lead to more credible counterfactuals since they estimate policy invariant parameters (e.g., user utility functions). Another advantage is that it allows researchers to incorporate domain knowledge in a principled manner, resulting in a statistical model that more accurately describes observations. These models are often estimated using ML techniques. One example of such work is Zhang et al. (2019), which models book readers' forward-looking utility-maximizing behavior using a forward-looking hidden Markov model (an ML technique for capturing sequential dependence) for improved targeting.

Here are some suggestions for writing effective papers using structural econometric models:

1. Briefly discuss, early in the paper, why one needs a structural econometric model (because they are often more complex than reduced form approaches), to answer the research question. It is helpful to spell it out in the study context.

2. Carefully present the underlying story of the model based on relevant theory and suggestive evidence from the dataset. Compare the proposed model with those based on alternative theories (using AIC/BIC, out-of-sample predictions, etc.) to show that it describes the phenomenon well.

3. As the main advantage of (and a motivation for) the structural econometric model is credible counterfactual analysis, ask meaningful business-relevant "what-if" questions and answer them using policy simulation based on the estimated model.

4. Here again, while the estimation of such models is challenging, this is not the key contribution. It is often better to use state-of-the-art estimation tools when available, while briefly discussing the unique estimation challenges.

In all three subtypes of studies in this type, ML methods are not contributions but aids to understand phenomena. Often such work displays innovative applications of ML techniques, illustrates practical challenges, suggests approaches to overcome them, and can spur new method development. These are valuable contributions to the literature. However, claiming new ML method contributions leads to confusion in the review process and eventually for the readers. In short, avoid confusing Type II ML papers with Type I.

Other Approaches

The influence of economic theory is noticeable in the three genres discussed above. This reflects the vibrant interface that currently exists between econometrics and ML and the strong economics-of-IS tradition in our field. But paradigms beyond economics can also be drawn upon. We have not seen as many of these submissions but we would be delighted to receive them.

For example, studies have used ML to extract psychometric or sociological variables followed by either quantitative or qualitative analysis. In these cases, it is still important to validate extracted variables, e.g., by recruiting experts to generate targets or evaluate subsets of predictions and presenting compelling evidence that the extracted variables capture the constructs motivated by the theory. When done well, such papers can provide novel and meaningful insights from much larger datasets than was previously possible (Ahmad et al. 2020; Miranda et al., in press).

Another example would be studies that identify novel data that help better predict outcomes of interest. Improvement in prediction can lead to an improved understanding of a phenomenon, especially for challenging problems. In such cases, researchers should surface the deeper underlying link between the new data and improvement in prediction to make an impactful contribution (e.g., Geva et al. 2017). Other examples include theoretical analyses, simulations, agent-based modeling, and various numerical methods. These approaches can help explore scenarios that can be difficult to collect data on, either because the research questions are forward looking or because it is difficult to find a setting with all possible scenarios of interest. When guided by appropriate theory and real-world observations, such research can provide important insights that would be difficult to obtain if we were only to rely on past observations.

Type III: ML in Complex Systems

Rather than study ML alone, Type III ML research focuses on learning how ML can contribute as part of a larger, complex system.[9] Complex systems research offers an alternative to *reductionist thinking*. As Mitchell (2009) notes: "the antireductionist catch-phrase, 'the whole is more than the sum of its parts,' takes on increasing significance as new sciences such as chaos, systems biology, evolutionary economics, and network theory move beyond reductionism to explain how complex behavior can arise from

[9] While "complex systems" can be defined in various ways, one influential definition was offered by the National Institute of Standards and Technology (NIST) as "large collections of interconnected components whose interactions lead to macroscopic behaviors in biological (e.g., ant colonies), physical (e.g., forest fires), social (e.g., economies) and information systems (e.g., compute clouds)" (see https://www.nist.gov/programs-projects/measurement-science-complex-information-systems).

large collections of simpler components." As just one sign of how far this line of thinking has come, the 2021 Nobel Prize in physics was awarded recently to three physicists for their collective work in using complex systems thinking to analyze problems ranging all the way from the smallest (e.g., atomic) to the largest scales (e.g., planetary systems, climate change).

IS researchers have long recommended using ideas from complex systems (e.g., Nolan and Wetherbe 1980; Merali 2006; and *MISQ*'s Special Issue on Complexity in 2020). Benbya et al. (2020, p. 3) offer an elegant example of a Google search to illustrate how a simple everyday activity depends on a complex sociotechnical maze: "while complexity in physical or social system is predominantly driven by either material operations or human agency, complexity in sociotechnical systems arises from the continuing and evolving entanglement of the social (human agency), the symbolic (symbol-based computation in digital technologies), and the material (physical artifacts that house or interact with computing machines)."

There is significant opportunity in viewing ML-driven information systems through this lens and posing research questions that aim to model and understand emergent behavior and/or purposefully design ML-driven information systems that *have* agency, while also being *aware* of human agency. In such a view of the world, some of the research questions might treat ML as a "black box" and study emergent behavior, along with efforts to design systems, while others might open up the black box, study and understand its properties in context (Saar-Tsechansky 2015), and explore how specific aspects of the ML model or algorithm themselves may contribute to emergent outcomes in a complex world—as well as how such aspects may be *designed* to build not just better ML models or algorithms but purposefully designed information systems too.

To illustrate the range of possible research contributions, consider the two seemingly simple scenarios reflected in Figures 2a and 2b. Figure 2a represents a scenario where the actions drawn from ML influence the environment, which in turn influences the data that is (subsequently, and continually) fed into the ML model. Explicitly modeling the environment and how the actions drawn from the ML model influence the data itself can be substantially complex and problem dependent but can reveal new avenues for research. Prawesh and Padmanabhan (2014, 2021), for example, have examined how news recommender systems are affected by such feedback processes. In their work, the recommender influences user behavior, which then affects the data that drives the recommender's subsequent performance. Other studies of this genre have shown how reinforcement learning can be used to model music playlist generation (Liebman et al. 2019) and how longitudinal dynamics of recommender systems can be understood using agent-based modeling (Zhang et al. 2020). Extending this view, the ML community has recently shown interest in a new type of learning problem—*performative prediction* (Perdomo et al. 2020)—to examine how decisions made from ML models can influence the outcomes they are designed to predict. This perspective gets exponentially more interesting when modeling multiple ML-driven agents (along with humans, as special types of agents), as shown in Figure 2b. For example, Malgonde et al. (2020) showed how multi-sided recommender systems could be studied from this perspective; one ML algorithm drives recommendation to one side of a digital platform (e.g., sellers), while another drives recommendations to another side (e.g., buyers).

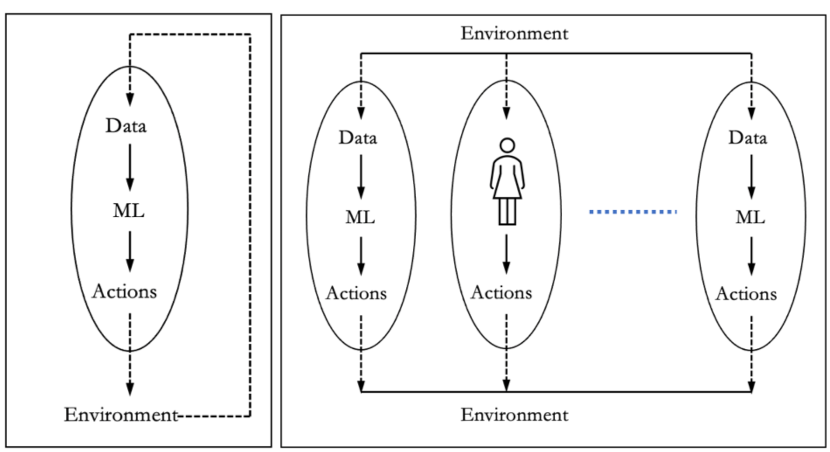

Figure 2. (a) Single-Agent Model (left), and (b) Multi-Agent Model (right)

From this multi-agent perspective, a combination of ML algorithms and humans with agency creates a rich range of possibilities and contributions to explore from a complex-systems perspective. Some possibilities include: (i) exploring how an ML-driven system, in conjunction with the environment, affects relevant emergent properties (such as fairness), and how it might trigger unexpected or unanticipated outcomes; and (ii) studying and designing specific types of agents in complex systems, i.e., with ML algorithms embodied in them and an environment (including human agency and interactions). We use the term "agents" broadly here, to include individuals, algorithms, information systems, or organizations; the environment surrounding them might itself be modeled as an agent or a set of agents within its own (sub)environment. While agent-based modeling is not necessarily required to explore this perspective (because carefully designed experiments may suffice), it is certainly a compelling paradigm to consider to integrate within ML work. Likewise, while authors do not need to use the design science paradigm to explore this perspective, we have found that the design science paradigm offers a natural and compelling framework for such work.

As a brief aside, we also note that entire organizational information systems could be explored from this perspective. Today we have data-driven ML models as part of organizational ERP systems, CRM systems, supply-chain systems, human resource management systems, cybersecurity systems, and communications and messaging systems such as email and collaboration tools. Understanding the role of ML in how these interact with each other within enterprises is important to advance our knowledge of how to better design organizational information systems in an ML-driven world.

When articulating contributions of work that fits into this framework, it is important to establish the significance of the issues studied and the novelty and significance of the contribution. Fairness of ML models in context, for instance, is clearly significant. However, novelty needs to be established clearly with respect to a growing body of knowledge. Similarly, the significance of the contribution will have to be articulated with respect to the body of knowledge that has accumulated in this area.

Authors should also ask if replacing "ML" with "Information Technology" broadly affects the message of the paper. If it remains much the same, the chances that this effect has already been shown in a more general (non-ML) IT context might be high. It is important for authors to reflect on whether showing an ML-specific effect in a context is a significant contribution when similar effects have been shown for other technologies in the same context. This criticism is particularly common when researchers coming from other disciplines submit ML-related research in IS journals without being aware of the history of IS research on those topics with earlier technologies. Considering what is unique about ML in this context compared to other types of technologies will help authors make better (ML) claims about their work. For instance, Berente et al. (2021) offer three ways—autonomy, learning, and inscrutability—in which AI is different. These (or similar) arguments can be used to make a convincing case that the wheel simply isn't being reinvented in a new context. The evaluation of ideas presented in this category could be analytical, computational (e.g., simulations), empirical, or experimental. Existing design science evaluation frameworks can also be used if it applies to what the authors are studying.

Some General Guidance

A significant part of this editorial has focused on helping identify specific types of ML-related contributions appropriate for *MISQ*. In addition, we would like to offer some general guidance for doing and writing/publishing ML-related work in Information Systems.

"Doing Research" Stage

1. Pinpoint your research problem. Real-world applications and datasets motivate many ML-papers. However exciting a new idea is, and however impactful the application is, researchers must take the time to consider and articulate the body of knowledge that their work builds on and contributes to. Put simply, what is the *research* problem in a particular body of literature? Being able to articulate this at multiple levels of abstraction will help authors to conduct their work, connect their research to (and differentiate their work from) others, choose the best outlets for their work, and identify suitable editors/reviewers to assess it.[10]

[10] Researchers might find that their work contributes to two different bodies of knowledge—one that is more methodological, and the other in the domain being studied. That is typically a sign that the authors have two papers rather than one. The authors must consider the relevant body of knowledge, and the contribution being made to that knowledge, separately for each paper, as the narrative will likely be distinct for each one.

2. Articulate your contribution. After identifying the target body of knowledge, authors must articulate precisely how they advance that knowledge and what they contribute. The contribution could fit into one of the three categories discussed in this editorial, or it could be something different.

3. Be pragmatically ambitious. Look for broad, impactful questions that our reference disciplines (including core ML research) are either not asking or are asking without having the right tools or lenses that IS researchers have to offer. Picking ML-related problems where IS researchers have a competitive advantage is important since CS departments and "big tech" industry labs often have large, well-funded teams and very rich data (particularly in industry). Researchers in those areas tend to publish in conferences with cycle times in months; in comparison, IS researchers might see cycle times in years due to the extensive multi-round review process that adds value but can reduce relevance, given the state of the art by the time of publication.

4. Respect the complexity of ML evaluation. Often in ML research, predictive accuracies alone in a task might be sufficient for publication. However, in IS, showing how these translate to business value is more compelling. Compared to the IS research community, the ML community has a better track record of working on problems with common benchmark datasets, which enable different algorithms to be compared to each other. In IS, we are yet to see such consensus, even on common problems (e.g., optimizing targeted advertisements). Until we have such common benchmarks, the onus will remain on the authors to convince a review team that a real business/social problem is being considered and that the evaluation does justice to the complexity of the problem. Of course, traditional means of evaluation (theoretical, analytical, computational, experimental) certainly also apply, and those should not be ignored. At the same time, we invite IS researchers to consider what it would take to construct benchmark evaluations in our field.

5. Be aware of ML research publication practices. ML research is *very* fast moving. By the time an ML-IS research project is under review, related papers may appear in the leading CS conferences. Reviewers will see them and could judge your work to be marginal in comparison. Careful problem definition (see #1 and #3 above) can address this to an extent. This journal is also open to authors submitting initial versions to the leading CS/ML conferences and then submitting extended versions for review. Presenting your ideas in these venues and incorporating feedback from there can strengthen your journal version.

In light of some confusion regarding prior conference papers in the publishing process, we would like to note the following. *MISQ* does have a policy that permits extensions of conference papers to be submitted to the journal. While we refer readers to the provenance statement[11] for detail, we wish to note two points relevant to our context here:

(1) If the paper is published in proceedings that are copyrighted (often the case for machine learning conferences), authors should ensure that there is substantive new material added to the paper before submitting to the journal. A rule of thumb is "30% additional material," which has some basis in policies from IEEE[12] and ACM[13]. These IEEE/ACM policies explicitly recognize that technical research often goes through different stages as part of the publication process; publishing in conferences followed by journals is recognized as an important part of the scientific process which deserves support. ACM even explicitly mentions that if 25% of substantive new material is added to a copyrighted (ACM) conference paper, the extended paper is in fact considered "new work" for which the author retains full copyright ownership.

(2) It is our expectation that the *entire* submitted paper will be holistically considered by review teams when assessing contributions, not just the "30% addition." Reviewers, and even editorial teams, have sometimes been unclear about this. We want authors to present their work at the best conferences, where possible, to get feedback to improve their work. The best ML conferences do have a high bar, and authors cannot "save" their best ideas for a journal version if they hope to get the paper in a highly selective ML conference. If the journal focuses on assessing only "incremental" contributions, then we end up in a catch-22: to produce the best work possible we want to encourage authors to try and send their ideas to the best conferences, but the best conferences won't accept ideas without substantial technical contribution, leaving authors in the hard spot of trying to demonstrate significant contributions in the incremental portion alone.

We absolutely want authors to substantively improve and extend their conference paper though, and the natural direction may be to add content that is more relevant from an IS/business school perspective (e.g., maybe new types of evaluation, broader

[11] https:// *MISQ*.umn.edu/provenance-service/
[12] https://www.ieee.org/publications/rights/section-822f.html
[13] https://authors.acm.org/author-services/author-rights

positioning, some algorithmic extensions where possible) and, in particular, enhancements or improvements based on feedback from the conference. In addition to the aforementioned "30% guideline," a rule of thumb might be for authors to think about what can be done to improve the paper with about four months of additional time to do so (approximate the time between a conference acceptance to when an author might submit to a journal after accounting for them presenting at the conference and getting feedback). More generally, whether you submit an earlier version of your paper to an ML conference or not, we strongly encourage you to workshop and refine your paper in any way you can before submitting it to *MISQ*. This has long been recommended practice (Weber 2002, p. ix). In our experience, one of the surest ways to get your paper rejected is for the editors and reviewers to be the first people to have read it. Please refine your paper in whatever way you can (e.g., workshops, seminars, and reviews by colleagues) before submitting it and explain how you did so in your cover letter.

"Writing Paper" Stage

1. Get the level of abstraction and the level of precision right. IS scholars vary in their knowledge of ML, but they invariably know (and are interested in) the broad themes and challenges. Linking a paper to these broad themes and challenges (especially in the introduction, discussion, and conclusion) improves its contribution (Rai 2017b). This does not mean, however, that authors should provide unnecessary background on everything in the paper. Review teams don't appreciate extensive coverage of common ML knowledge (a common mistake by new authors). Providing proper references is sufficient. Hence, in the main sections of the paper, write to a more informed (ML) review team, while providing useful references along the way as needed.

2. Focus on the "core" new ideas. Authors invariably find every part of their paper to be important, but they are responsible for finding the core idea and communicating that clearly. If the core idea depends on other components that also need to be described, do so, but clearly point out which components are novel so that review teams can directly focus on those elements.

3. Clearly articulate contributions early in the paper (Hevner et al. 2004; Saar-Tsechansky 2015). It is surprising how many submissions still do not explicitly note the specific contributions being made. This should not be left to the editors and reviewers to figure out because authors are in the best position to pinpoint what their novel contribution is. "Over-claiming" can hurt since review teams are quick to find other work that may have already shown some of the claimed contributions in related contexts. Being specific and concise when articulating the main contributions of the work is a better course of action. A related issue arises for certain types of ML papers when a new method is presented. Often the new method uses many ideas from existing approaches, and when the authors articulate all of this in a manner that suggests that the *entire* approach is novel, review teams are naturally unimpressed. Even when presenting the (new) method, underscoring which parts represent the main contribution and which ones come from prior knowledge is important. If the *synthesis* of preexisting approaches is novel and is the key contribution (again, the authors are in the best position to determine this), that synthesis should be highlighted while acknowledging the existing components so that there is no confusion regarding the contribution claim.

4. Clearly position the contribution with respect to the proper body of knowledge being advanced. If authors do not do this clearly, it may suggest to a review team that they themselves do not yet know which body of knowledge the paper is advancing and how. For application-inspired ML work, the body of knowledge might be what is known at the intersection of ML and the context under consideration. Authors will benefit greatly in the review process if they provide a thorough review of related work and place their new idea and its contributions in that light.

5. Get early feedback. ML is a very active area of research. Getting feedback on early drafts from other researchers and/or from presenting the ideas in mainstream ML/IS conferences will help to both gauge interest and develop the ideas.

Conclusion and the Road Ahead

We are excited about how ML and AI can transform the world and welcome the unique opportunity in front of us as IS researchers to shape this transformation. Technology can evolve in unpredictable ways; not surprisingly, we are seeing signs that ML and AI might have such aspects to them as well. Given our backgrounds at the intersection of computational and social sciences, IS researchers will most likely ask and answer ML-related questions in unique ways that differ from the questions asked by our colleagues in the computer science community, with its "big tech focus," and from our colleagues in other applied disciplines. We are seeing significant interest in ML research in the IS field and are receiving many ML-related submissions to *MISQ*. However, the lack of a clear "template" of how to do, write-up, and position ML-related work in IS has challenged both authors and review teams. It was against this backdrop that we launched two masterclasses in ML for our editors and authors, and followed those up with this editorial to help authors shape their ideas in ways that can amplify their impact while reducing some of the uncertainties in the review process. We hope you find our ideas to be helpful and we look forward to seeing your papers submitted to *MISQ*.

Acknowledgments

We thank the 500+ participants in two masterclasses aimed at authors for their valuable questions and comments. We also thank the entire *MISQ* editorial board for their feedback and comments from another masterclass aimed at their role. We also thank the colleagues who provided detailed and thoughtful comments on a version of this editorial, including Gedas Adomavicius, Nick Berente, Tomer Geva, Bin Gu, Wen Hua, Yan Leng, Jan Recker, Maytal Saar-Tsechansky, Shazia Sadiq, Radhika Santhanam, Sumit Sarkar, Konstantina Valogianni, Tong Wang, Yingfei Wang, Sen Wang, and Kunpeng Zhang. We were able to incorporate many of their suggestions directly here, and hope to use other forums to continue this conversation with our "ML in IS" community of researchers.

References

Abbasi, A., and Chen, H. 2008. "CyberGate: A Design Framework and System for Text Analysis of Computer-Mediated Communication," *MIS Quarterly* (32:4), pp. 811-837.

Ahmad, F., Abbasi, A., Li, J., Dobolyi, D. G., Netemeyer, R. G., Clifford, G. D., and Chen, H. 2020. "A Deep Learning Architecture for Psychometric Natural Language Processing," *ACM Transactions on Information Systems* (38:1), Article 6.

Adomavicius, G., and Tuzhilin, A. 2005. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering* (17:6), pp. 734-749.

Adomavicius, G., Tuzhilin, A., and Zheng, R. 2011. "REQUEST: A Query Language for Customizing Recommendations," *Information Systems Research* (22:1), pp. 99-117.

Angrist, J. D., and Pischke, J.-S. 2008. *Mostly Harmless Econometrics*, Princeton University Press.

Archak, N., Ghose, A., and Ipeirotis, P. G. 2011. "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science* (57:8), pp. 1485-1509.

Belloni, A., Chernozhukov, V., and Hansen, C. 2013. "Inference on Treatment Effects after Selection among High-Dimensional Controls," *The Review of Economic Studies* (81:2), pp. 608-650.

Benbya, H., Nan, N., Tanriverdi, H., and Yoo, Y. 2020. "Complexity and Information Systems Research in the Emerging Digital World," *MIS Quarterly* (44:1), pp. 1-17.

Berente, N., Gu, B., Recker, J., & Santhanam, R. 2021. Managing Artificial Intelligence. *MIS Quarterly* (45:3), pp. 1433-1450.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. 2017. "Geometric Deep Learning: Going beyond Euclidean Data," *IEEE Signal Processing Magazine* (34:4), pp. 18-42.

Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., and Walsh, A. 2018. "Machine Learning for Molecular and Materials Science," *Nature* (559:7715), pp. 547-555.

Caruana, R. 1997. "Multitask Learning," *Machine Learning* (28:1), pp. 41-75.

Caruana, R., Lundberg, S., Ribeiro, M. T., Nori, H., and Jenkins, S. 2020. "Intelligible and Explainable Machine Learning: Best Practices and Practical Challenges," *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3511-3512.

Chen, W., Wang, Y., and Yang, S. 2009. "Efficient influence maximization in social networks," *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 199-208.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal* (21:1), pp. C1-C68.

Cohn, D. A., Ghahramani, Z., and Jordan, M. I. 1996. "Active Learning with Statistical Models," *Journal of Artificial Intelligence Research* (4), pp. 129-145.

Domingos, P., and Richardson, M. 2001. "Mining the Network Value of Customers," *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 57-66.

Fang, X., Liu Sheng, O. R., and Goes, P. 2013. "When is the Right Time to Refresh Knowledge Discovered from Data?" *Operations Research* (61:1), pp. 32-44.

Fang, X., and Hu, P. J. 2018. "Top Persuader Prediction for Social Networks," *MIS Quarterly* (42:1), pp. 63-82.

Fang, X., Gao, Y., and Hu. P. J. 2021. "A Prescriptive Analytics Method for Cost Reduction in Clinical Decision Making," *MIS Quarterly*, (45:1), pp. 83-115.
Floridi, L., and Chiriatti, M. 2020. "GPT-3: Its Nature, Scope, Limits, and Consequences," *Minds and Machines* (30:4), pp. 681-694.
Geva, T., Oestreicher-Singer, G., Efron N., and Shimshoni, Y. 2017. "Using Forum and Search Data for Sales Prediction of High-Involvement Projects," *MIS Quarterly* (41:1), pp. 65-82.
Gregor, S., and Hevner, A. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly*, (37:2), pp. 337-355.
Gorgoglione, M., Panniello, U., and Tuzhilin, A. 2016. "In CARSs We Trust: How Context-Aware Recommendations Affect Customers' Trust and Other Business Performance Measures of Recommender Systems," *Information Systems Research*, (27:1), pp. 182-196.
Gupta, A. 2018. "Traits of Successful Research Contributions for Publication in ISR: Some Thoughts for Authors and Reviewers," *Information Systems Research* (29:4), pp. 779-786.
Hartford, J., Lewis, G., Leyton-Brown, K., and Taddy, M. 2017. "Deep IV: A Flexible Approach for Counterfactual Prediction," in *Proceedings of the 34th International Conference on Machine Learning*, pp. 1414-1423.
He, J., Fang, X., Liu, H., Li, X. 2019. "Mobile App Recommendation: An Involvement-Enhanced Approach," *MIS Quarterly* (43:3), pp. 827-849.
Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., and Meger, D. 2018. "Deep Reinforcement Learning that Matters," in *Proceedings of the AAAI Conference on Artificial Intelligence*.
Hendershott, T., Zhang, M. X., Zhao, J. L., and Zheng, E. 2017. "Call for Papers—Special Issue of Information Systems Research Fintech: Innovating the Financial Industry through Emerging Information Technologies," *Information Systems Research*, (28:4), pp. 885-886.
Hevner, A., March, S., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-105.
Jain, H., Padmanabhan, B., Pavlou, P. A., and Raghu, T. S. 2021. "Editorial for the Special Section on Humans, Algorithms, and Augmented Intelligence: The Future of Work, Organizations, and Society," *Information Systems Research* (32:3), pp. 675-687.
Johnstone, I. M., and Titterington, D. M. 2009. "Statistical Challenges of High-Dimensional Data." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (367), pp. 4237-4253.
Kong, D., and Saar-Tsechansky, M. 2014. "Collaborative Information Acquisition for Data-Driven Decisions," *Machine Learning*, (95:1), pp. 71-86.
LeCun, Y., Bengio, Y., and Hinton, G. 2015. "Deep Learning," *Nature*, (521:7553), pp. 436-444.
LeDell, E., and Poirier, S. 2020. "H2o AutoML: Scalable Automatic Machine Learning," in *Proceedings of the AutoML Workshop at ICML*.
Lee, A. S. 1999. "Inaugural Editor's Comments," *MIS Quarterly*, (23:1), pp. v-xi.
Li, T., Sahu, A. K., Talwalkar, A., and Smith, V. 2020. "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Processing Magazine* (37:3), pp. 50-60.
Li, Z., Fang, X., Bai, X., and Liu Sheng, O. R. 2017. "Utility-Based Link Recommendation for Online Social Networks," *Management Science* (63:6), pp. 1938-1952.
Liang, T.P. 1988. "Model Management for Group Decision Support," *MIS Quarterly* (12:4) pp. 667-680.
Liebman, E., Saar-Tsechansky, M., and Stone, P. 2019. "The Right Music at the Right Time: Adaptive Personalized Playlists Based on Sequence Modeling," *MIS Quarterly* (43:3), pp. 765-786.
Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. 2021. "Self-Supervised Learning: Generative or Contrastive," *IEEE Transactions on Knowledge and Data Engineering*, doi: 10.1109/TKDE.2021.3090866.
Lukyanenko, R., Parsons, J., Wiersma, Y. F., and Maddah, M. 2019. "Expecting the Unexpected: Effects of Data Collection Design Choices on the Quality of Crowdsourced User-Generated Content," *MIS Quarterly* (43:2), pp. 623-648.
Malgonde, O., Zhang, H., Padmanabhan, B., and Limayem, M. 2020. "Taming Complexity in Search Matching: Two-Sided Recommender Systems on Digital Platforms," *MIS Quarterly*, (44:1) pp. 49-84
Menon, S., and Sarkar, S. 2016. "Privacy and Big Data: Scalable Approaches to Sanitize Large Transactional Databases for Sharing," *MIS Quarterly* (40:4), pp. 963-981.
Merali, Y. 2006. "Complexity and Information Systems: The Emergent Domain," *Journal of Information Technology* (21:4), pp. 216-228.
Meyer, G., Adomavicius, G., Johnson, P. E., Elidrisi, M., Rush, W. A., Sperl-Hillen, J. M., and O'Connor, P. J. 2014. "A Machine Learning Approach to Improving Dynamic Decision Making," *Information Systems Research* (25:2), pp. 239-263.

Miranda, S. M., Wang, D. D., and Tian, C. A. (in press). "Discursive Fields and the Diversity-Coherence Innovation Paradox: An Ecological Perspective on the Blockchain Community Discourse," *MIS Quarterly*.

Mitchell, M. 2009. *Complexity: A Guided Tour*. Oxford University Press.

Mitchell, T. M. 1997. *Machine Learning* (1st. ed.), McGraw-Hill.

Newey, W. K., and Powell, J. L. 2003. "Instrumental Variable Estimation of Nonparametric Models," *Econometrica* (71:5), pp. 1565-1578.

Niederman, F., and March, S. 2012. "Design Science and the Accumulation Of Knowledge in the Information Systems Discipline," *ACM Transactions on Management Information Systems* (3:1), Article 1.

Nolan, R. L and Wetherbe, J. C. (1980) Toward a Comprehensive Framework for MIS Research, *MIS Quarterly* (4:2) pp. 1-19.

Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., and Taylor, J. 2019. "Industry-Scale Knowledge Graphs: Lessons and Challenges," *Communications of the ACM*, (62:8), pp. 36-43.

Oprescu, M., Syrgkanis, V., and Wu, Z. S. 2019. "Orthogonal Random Forest for Causal Inference," in *Proceedings of the 36th International Conference on Machine Learning*, pp. 4932-4941.

Padmanabhan, B., and Barfar, A. 2021. "Learning Individual Preferences from Aggregate Data: A Genetic Algorithm for Discovering Baskets of Television Shows with Affinities to Political and Social Interests," *Expert Systems with Applications* (168), Article 114184.

Padmanabhan, B., and Tuzhilin, A. 2003. "On the Use of Optimization for Data Mining: Theoretical Interactions and eCRM Opportunities," *Management Science* (49:10), pp. 1327-1343.

Pearl, J. 2019. "The Seven Tools of Causal Inference, with Reflections on Machine Learning," *Communications of the ACM* (62:3), pp. 54-60.

Perdomo, J., Zrnic, T., Mendler-Dünner, C., and Hardt, M. 2020. "Performative Prediction," *International Conference on Machine Learning* (pp. 7599-7609).

Prawesh, S., and Padmanabhan, B. 2014. "The 'Most Popular News' Recommender: Count Amplification and Manipulation Resistance," *Information Systems Research* (25:3), pp. 569-589.

Prawesh, S., and Padmanabhan, B. 2021. "A Complex Systems Perspective of News Recommender Systems: Guiding Emergent Outcomes with Feedback Models," *PLOS One* (16:1), Article e0245096.

Purao, S., Storey, V. C., and Han, T. 2003. "Improving Analysis Pattern Reuse in Conceptual Design: Augmenting Automated Processes with Supervised Learning," *Information Systems Research* (14:3), pp. 269-290.

Rai, A. 2017a. "Diversity of Design Science Research," *MIS Quarterly* (41:1), pp. iii-xviii.

Rai, A. 2017b. "Editor's Comments: Seeing the Forest for the Trees," *MIS Quarterly* (41:4), pp. iii-vii.

Ram, S., and Goes, P. 2021. "Focusing on Programmatic High Impact Information Systems Research, Not Theory, to Address Grand Challenges," *MIS Quarterly* (45:1), pp. 479-483.

Saar-Tsechansky, M. 2015. "Editor's Comments: The Business of Business Data Science in IS Journals," *MIS Quarterly* (39:4), pp. iii-vi.

Saar-Tsechansky, M., and Provost, F. 2007. "Decision-Centric Active Learning of Binary-Outcome Models," *Information Systems Research* (18:1), pp. 4-22.

Sadiq, S., Aryani, A., Demartini, G., Hua, W., Indulska, M., Burton-Jones, A., Khosravi, H. et al. 2022. "Information Resilience: The Nexus of Responsible and Agile Approaches to Information Use," *The VLDB Journal*, doi: 10.1007/s00778-021-00720-2.

Samtani, S., Chai, Y., and Chen, H. (in press). "Linking Exploits from the Dark Web to Known Vulnerabilities for Proactive Cyber Threat Intelligence: An Attention-Based Deep Structured Semantic Model Approach", *MIS Quarterly*.

Sahoo, N., Krishnan, R., Duncan, G., and Callan, J. 2012. "Research Note—The Halo Effect in Multicomponent Ratings and its Implications for Recommender Systems: The Case Of Yahoo! Movies," *Information Systems Research* (23:1), pp. 231-246.

Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., and Lee, K. C. 2020. "Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach," *MIS Quarterly* (44:4), pp. 1459-1492.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., and Dieleman, S. 2016. "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature* (529:7587), pp.484-489.

Simchi-Levi, D. 2020. "From the Editor," *Management Science* (66:1), pp. 1-4.

Shmueli, G. 2010. "To Explain or to Predict?," Statistical Science (25:3), pp. 289-310.

Stoyanovich, J., Howe, B., & Jagadish, H. V. 2020. "Responsible Data Management," *Proceedings of the VLDB Endowment*, (13:12), pp. 3474-3488.

Sutton, R. S., and Barto, A. G. 2018. *Reinforcement Learning: An Introduction.* MIT Press.

Tomašev, N., Glorot, X., Rae, J. W., Zielinski, M., Askham, H., Saraiva, A., Mottram, A., Meyer, C., Ravuri, S., Protsyuk, I., and Connell, A. 2019. "A Clinically Applicable Approach to Continuous Prediction of Future Acute Kidney Injury," *Nature* (572:7767), pp. 116-119.

Torrey, L., and Shavlik, J. 2010. "Transfer Learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, Olivas, E.S., Guerrero, J. D. M., Martinez-Sober, M., Magdalena-Benedito, J. R., and Lopez, A. J. S. (Eds.), IGI Global, pp. 242-264.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. 2017. "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000-6010.

Wager, S., and Athey, S. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests," *Journal of the American Statistical Association* (113:523), pp. 1228-1242.

Wang, R.Y., and Strong, D. M. 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems* (12:4), pp. 5-33.

Weber, R. 2002. "Editor's Comments: Retrospection: The MIS Quarterly's Review Processes: 1995-2001," *MIS Quarterly* (26:2), pp. v-xi.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. 2020. "A Comprehensive Survey on Graph Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems* (32:1), pp. 4-24.

Xu, D., Hu, P., and Fang, X. 2021. "A Deep Learning-Based Imputation Method to Enhance Crowdsourced Data for Online Business Directory Platforms," Working Paper.

Xu, L., Duan, J. A., and Whinston, A. 2014. "Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion," *Management Science* (60:6), pp. 1392-1412.

Yang, M., Adomavicius, G., Burtch, G., and Ren, Y. 2018. "Mind the Gap: Accounting for Measurement Error and Misclassification in Variables Generated via Data Mining," *Information Systems Research* (29:1), pp. 4-24.

Yeh, R. A., Chen, C., Yian Lim, T., Schwing, A. G., Hasegawa-Johnson, M., and Do, M. N. 2017. "Semantic Image Inpainting with Deep Generative Models," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5485-5493.

Zhang, J., Adomavicius, G., Gupta, A., and Ketter, W. 2020. "Consumption and Performance: Understanding Longitudinal Dynamics of Recommender Systems via an Agent-Based Simulation Framework," *Information Systems Research* (31:1), pp. 76-101.

Zhang, S., Lee, D., Singh, P. V., and Srinivasan K. (in press). "What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features," *Management Science*.

Zhang, Y., Li, B., Luo, X., and Wang, X. 2019. "Personalized Mobile Targeting with User Engagement Stages: Combining a Structural Hidden Markov Model and Field Experiment," *Information Systems Research* (30:3), pp. 787-804.

Zhu, X., and Goldberg, A. B. 2009. "Introduction to semi-supervised learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning* (3:1), pp. 1-130.

Editorial Board Changes for 2022

The following editors completed their terms on the *MISQ* editorial board in December 2021. We thank them for their dedicated service and their contributions to authors and the field.

Associate Editors

Ofer Arazy, University of Haifa
Nicholas Berente, University of Notre Dame
Panos Constantinides, University of Manchester
John D'Arcy, University of Delaware
Xiao Fang, University of Delaware
Nirup Menon, George Mason University
Ilan Oshri, University of Auckland
Nilesh Saraf, Simon Fraser University
Ali Tafti, University of Illinois Chicago
Ryan Wright, University of Virginia
Bo Sophia Xiao, University of Hawai'i at Mānoa
Heng Xu, American University
Xiaojun Zhang, Hong Kong University of Science and Technology

Senior Editors

Gediminas Adomavicius, University of Minnesota
Dennis Galletta, University of Pittsburgh
Saonee Sarker, Lund University[14]

The following editors started their terms on the board in January 2022. We are thrilled to have such exceptional scholars join our team. We know they look forward to serving.

Associate Editors

Hanna Halaburda, New York University
Cheng Suang Heng, National University of Singapore
Marleen Huysman, Vrije Universiteit Amsterdam
Hee-Woong Kim, Yonsei University
Nishtha Langer, Rensselaer Polytechnic Institute
Jingjing Li, University of Virginia
YoungKi Park, George Washington University
Lisen Selander, University of Gothenburg
Ping Wang, University of Maryland, College Park
Jaime Windeler, University of Cincinnati
Cheng Yi, Tsinghua University

Senior Editors

Panos Constantinides, University of Manchester
John D'Arcy, University of Delaware
Maytal Saar-Tsechansky, University of Texas, Austin
Ann Majchrzak, University of Southern California

[14] Saonee continues to serve on the board as our inaugural director of diversity, equity, and inclusion.