WARWICK
THE UNIVERSITY OF WARWICK

Manuscript version: Author's Accepted Manuscript
The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:
http://wrap.warwick.ac.uk/134220

How to cite:
Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

THE ROLE OF ARTIFICIAL INTELLIGENCE AND DATA NETWORK EFFECTS FOR CREATING USER VALUE

(post-print version, forthcoming in *Academy of Management Review*)

Robert Wayne Gregory
University of Virginia, rg7cv@virginia.edu

Ola Henfridsson
University of Miami, ohenfridsson@miami.edu

Evgeny Kaganer
Moscow School of Management, SKOLKOVO, evgeny_kaganer@skolkovo.ru

Harris Kyriakou
IESE Business School, hkyriakou@iese.edu

Acknowledgements:

We sincerely thank associate editor Allan Afuah and three anonymous reviewers for their constructive guidance on developing the paper. We are also grateful for the opportunity to participate in the AMR Theory Development Hackathon held at IESE Business School in Barcelona in February 2019.

THE ROLE OF ARTIFICIAL INTELLIGENCE AND DATA NETWORK EFFECTS FOR CREATING USER VALUE

Some of the world's most profitable firms own platforms that exhibit network effects. A platform exhibits network effects if the more that people use it, the more valuable it becomes to each user. Theorizing about the value perceived by users of a platform that exhibits network effects has traditionally focused on direct and indirect network effects. In this paper, we theorize about a new category of network effects—data network effects—that has emerged from advances in artificial intelligence (AI) and the growing availability of data. A platform exhibits data network effects if the more that the platform learns from the data it collects on users, the more valuable the platform becomes to each user. We argue that there is a positive direct relationship between the AI capability of a platform and the value perceived in the platform by its users—a relationship that is moderated by platform legitimation, data stewardship and user-centric design.

INTRODUCTION

Network effects make crucial contributions to the value that users perceive in the products, services, or platforms of some of the world's most valuable firms (e.g., Apple, Microsoft, Facebook). A platform or one of its products or services exhibits network effects if the more that people use it, the more valuable it becomes to each user (Church & Gandal, 1992; Farrell & Saloner, 1985, 1986; Katz & Shapiro, 1985, 1986, 1992; Sheremata, 2004; Suarez, 2005). For example, a social network such as Facebook exhibits network effects because the more that people use it, the more valuable it becomes to each user since more users mean more people to interact with (Afuah, 2013; Van Alstyne, Parker, & Choudary, 2016). Because of the immense impact that network effects can have on the value that users perceive in a platform, many scholars have theorized about their nature and consequences for user value (Cennamo & Santalo, 2013; Eisenmann, Parker, & Van Alstyne, 2011; Gawer, 2009; Majumdar & Venkataraman, 1998; Mcintyre & Srinivasan, 2017; Mcintyre, Srinivasan, Afuah, Gawer, & Krestschmer, 2020; Parker & Van Alstyne, 2005; Priem, Butler, & Li, 2013; Rochet & Tirole, 2003).

To date, research has focused on two categories of network effects: direct network effects and indirect network effects (Clements, 2004; Mcintyre & Srinivasan, 2017). In the case of direct network effects, the value that users derive from a network comes from users being able to interact directly with each other (Rochet & Tirole, 2003; Zhu & Iansiti, 2012). For example, network effects on social media platforms primarily stem from users interacting directly with each other. In the case of indirect network effects, the more people that use a product, the higher is the likelihood of increased availability and variety of complements of the product, thereby increasing the value of the product to each user (Boudreau, 2012; Church, Gandal, Krause, & Canada, 2008; Clements & Ohashi, 2005). For example, the more users that are attracted to a mobile ecosystem, the greater are the incentives for development and thus the diversity of apps, resulting in more perceived user value of products within that mobile ecosystem. In sum, extant research has effectively explored the impact of network effects on the value perceived by users in terms of both direct and indirect network effects.

However, little attention has been paid to *data network effects* as an emerging category of network effects. A platform exhibits data network effects if, the more that the platform learns from the data it collects on users, the more valuable the platform becomes to each user. For example, the more that Google learns about users and the searches that they conduct, the more it can individualize the experience, making the search engine more valuable to each user. Similarly, the more that Tesla optimizes its self-driving algorithms by feeding them with billions of miles worth of driving data it gathers from in-car sensors, cameras, and radar units, the greater is the perceived value of Tesla cars.

In this paper, we explore the role of artificial intelligence (AI) and data network effects for creating user value, especially in the context of multi-sided platforms. The starting point is the observation that the value each user perceives depends on the scale of data-driven learning and improvements realized with AI. Such learning and improvements typically rely on faster and better predictions through applications of machine learning grounded in data (Agrawal, Gans, & Goldfarb, 2018; Samuel, 1959). For example, music streaming services use machine learning techniques to continuously learn about users' listening preferences and improve their recommendation engine, making the platform more valuable to each user.

Assuming direct connections and multi-sided exchange between users or user groups (e.g., Farrell & Saloner, 1985; Katz & Shapiro, 1985), prior network effects literature cannot readily explain why data-driven learning and improvements on a platform contributes to user value through data network effects. In this research, we therefore examine the role of AI and the "computer in the middle of every transaction" (Varian, 2014, p. 1) to address the following research question: *how is value for each user of the platform created from data with AI?*

We develop a model of data network effects that complements and extends existing network effects theory (Cennamo & Santalo, 2013; Church et al., 2008; Farrell & Saloner, 1986; Katz & Shapiro, 1986; Parker & Van Alstyne, 2005; Rochet & Tirole, 2003). Based on the premise of same-side or multi-sided exchange among users, network effects theory posits that a growing network of interconnected users gives rise to network externalities, where a user's utility of a platform is a function of the total number of users (Katz & Shapiro, 1985). In this paper, we propose that platform AI capability, i.e., the ability of a platform to learn from data to continuously improve its products or services for each user, gives rise to new platform externalities, where a user's utility of a platform is a function of the scale of data-driven learning and improvements realized with AI. These improvements manifest in greater product functionality, platform quality, and experience for each user (Cennamo & Santalo, 2013; Cennamo & Santaló, 2018; Himan, 2002; Mcintyre & Srinivasan, 2017; Zhu & Iansiti, 2012; Zhu & Liu, 2018). Our model of data network effects explains this novel phenomenon.

CONCEPTUAL BACKGROUND

Before presenting the model, we first summarize the background information about artificial intelligence and network effects that is needed to understand the causal arguments of the model.

Artificial Intelligence

Brian Arthur (2009) proposes three principles through which we can understand advanced technologies such as AI: combination, recursiveness, and phenomena. First, while AI pioneers in the 1950s such as John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon projected that "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" (McCorduck, 2004), today's application of AI exhibits a more modest ambition by *combining* technologies in particular functional domains (Raisch & Krakowski, 2020). Advances in machine learning offer a novel approach to specific decision-making tasks and business problems (Finlay, 2017). In turn, such machine learning draws on the dramatically improved performance-price ratio of computer processing technology, data storage and management, and network technologies (Agrawal et al., 2018; Yoo, Henfridsson, & Lyytinen, 2010). In combination, these technologies make AI an important tool for enabling platforms, products, or services to generate user value. For example, navigation services leverage data collected about users to offer dynamic turn-by-turn navigation based on continuously improved predictions of traffic situations. The perceived value for each user increases as predictions benefit from the increasing processing and networking power of computing devices such as smartphones.

Second, AI applications exhibit a modular architecture (Garud & Karnøe, 2003; Schilling, 2000) making up a complex network of technologies where each technology is developed independently with its own set of design objectives. This creates *recursiveness* that influences the perceived user value of the service employing AI. AI consists of technologies, which in turn consist of technologies. As improvements in one piece of technology are accomplished, this may conflict with the design objectives of another piece of technology. For instance, consider how the improvements in the recommendation engine of a service may improve its convenience, personalization, and ease of use. However, the improvements in the engine may cause privacy concerns and require improvements in privacy protection and the cybersecurity technology used. In fact, in the wake of recursiveness, firms leveraging AI may face scrutiny from key stakeholder audiences regarding their collection and use of personal data, the lack of transparency in their decisions (e.g., automated loan decision-making), and how they deal with the errors stemming from biases oftentimes inherited when algorithms use data collected on users (Ahsen, Ayvaci, & Raghunathan, 2019).

Finally, resonating with Arthur's third principle, that of *phenomena*, today's AI has data-driven learning at its center (Meyer et al., 2014). For example, rather than programming explicit rules for recognizing a cat or a dog, neural networks (a form of machine learning algorithms) are capable of teaching themselves classification if trained with a prelabeled dataset. In practice, training machine learning algorithms involves much tinkering and experimentation, iteratively learning from data to detect patterns and predict outcomes faster and more accurately (Agrawal et al., 2018). In this regard, the value of AI technologies is based on the existence of big data (McAfee & Brynjolfsson, 2012; Varian, 2014). Big data refers to very large volumes of data, the ability to process and transmit that data at a high velocity, the existence of an increasing variety of data sources including social networks, mobile devices, connected things, and open data (weather, traffic, maps, etc.), and the challenge of ensuring veracity so that data sources truly represent reality (Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2016). The volume, variety, and veracity of data make important contributions to predictive model development from which users will more likely benefit. In addition, trained prediction models enable data-driven products and services to continuously learn and improve on the basis of feedback data from users who share a variety of personal data at a high velocity.

Network Effects

The concept of network effects is predicated on the notion that network externalities give rise to value creation through direct connections or multi-sided exchange among individual users or different groups of users on opposite sides of the market. As outlined by Katz & Shapiro (1985), the utility that a user derives from a platform is a function of the total number of users, as the scale of the network gives rise to consumption or network externalities. Using network size as the main determinant of user value (Parker & Van Alstyne, 2018; Suarez, 2005), network effects theory examines how increases in the network size of one user group may produce a virtuous cycle with increases in the network size of either the same user group (direct network effects) or another user group, providing complements to the platform (indirect network effects) (Church & Gandal, 1992; Church et al., 2008; Katz & Shapiro, 1992; Rochet & Tirole, 2003, 2006; Schilling, 2002).

However, the broad adoption and diffusion of AI on today's platforms warrants another look at network effects. In particular, it should be noted that "a computer in the middle of every transaction" (Varian, 2014, p. 1) does not merely provide connectivity and possibilities for exchange among users. It also gives rise to new data networks that platform companies explore and exploit with the help of AI. By means of automation or augmentation (Raisch & Krakowski, 2020), AI enables significant scaling of the learning from the data collected on users as they leave digital traces of interconnections with things, people, and organizations in their daily life. Whether or not these new processes of data-driven value creation and capture have a positive or negative effect on the perceived value for each user of the platform (Stucke & Ezrachi, 2016, 2018; Tucker, 2019), they give rise to new platform externalities (Himan, 2002) underlying the concept of data network effects. The utility that a user derives from a platform is then a function of the scale of data-driven learning and improvements realized with AI. The resulting data network effect, in terms of the increase in user value, may manifest in superior (inferior) functionalities of the products delivered through the platform, a more (less) personalized and meaningful experience for each user, or other aspects of platform quality (see e.g., Cennamo & Santalo, 2013; Cennamo & Santaló, 2018; Himan, 2002; Mcintyre & Srinivasan, 2017; Zhu & Iansiti, 2012; Zhu & Liu, 2018).

Recent work by Afuah (2013) provides a good starting point for exploring these data network effects because of the departure from focusing on network size as the only determinant of user value. In particular, Afuah (2013) proposes network structure and conduct as additional factors that contribute to user value on multisided platforms. For instance, network structure may vary in terms of how it contributes to user value through the degree of transaction feasibility. The feasibility of transactions depends not only on the existence of connectivity among users but also on the availability of data and useful information on the network to achieve the best possible matches between supply and demand and to enable each individual user to make more informed decisions on entering and executing transactions (Chen & Horton, 2016). Uber, for example, uses machine learning algorithms on its platform to analyze data collected on each user in real-time and to improve both the algorithmic matching as well as the information and experience offered to each user who decides to engage in exchanges among riders and drivers. As this example illustrates, the feasibility of transactions also depends on the extent to which each user is actively engaged in using the platform or its products and services. To this end, Uber is known for using techniques of behavioral nudging to inform and engage users with the help of push notifications and messages providing intelligent recommendations that adapt to changing contextual and situational circumstances (Rosenblat, 2018).

The growing influence of learning from data on the network in the era of AI also becomes evident when considering network conduct (Afuah, 2013), another factor contributing to user value. For instance, not all users of the network are necessarily rational and have identical information about each other and the possible transactions. This may result in opportunistic behavior that makes the platform less valuable, on average, to users. For example, some Uber drivers may try to game the system by going offline to avoid fulfilling passenger requests that they find less lucrative than those received on an alternative ride hailing platform that they use in parallel. By learning from data collected on each user through the use of machine learning algorithms on the platform, Uber tries to prevent what it views as fraudulent behavior, an instance of opportunistic behavior. On the other hand, users of a platform may earn a reputation for being trustworthy, dependable, and honest, which may positively impact the contribution of network conduct to user value because this reputation serves as a signal to other users and motivates exchange. In the example of Uber, drivers and passengers rate each other. A positive five-star rating helps a driver obtain repeated jobs, while lower ratings create an important barrier for deriving value from the platform. Finally, the perception of trust in the platform, the object of exchange or the exchange partners themselves may also play an important role in users engaging on the platform and obtaining benefits. In the case of Uber, the behavior and exchange relationship between passengers and drivers is governed by the use of machine learning algorithms on the platform (Rosenblat, 2018).

In sum, the underlying mechanism of how data network effects contribute to user value influences the network by increasing the scale of learning from the data collected on users through the use of AI. To substantiate this claim, we develop a framework addressing our research question of how the value for each user of the platform is created from data through AI.

A FRAMEWORK FOR EXPLORING THE ROLE OF ARTIFICIAL INTELLIGENCE AND DATA NETWORK EFFECTS FOR CREATING USER VALUE

Figure 1 shows our framework for explaining the role of AI and data network effects for creating user value, defined as the value that users perceive in the platform (e.g., Facebook) or its products and services (e.g., News Feed, Pages). The data network effects themselves are manifested in the positive direct relationship between the AI capability of a platform and the value of the platform as perceived by its users—a relationship that is moderated by platform legitimation, data stewardship and user-centric design.

Insert Figure 1 about here

The framework is based on the following set of assumptions:

(1) The "computer in the middle of every transaction" (Varian, 2014, p. 1) turns AI-enabled platforms into flexible infrastructures that are *capable of learning* (Assumption 1). For example, in addition to employing plenty of human labor, social media sites such as Facebook and Inke, one of the largest Chinese live-streaming companies, use machine learning algorithms to help moderate (e.g., find and remove) toxic content, including spam, hate speech, nudity, violence, and terrorist propaganda.

(2) The strategic role of machine learning in today's platforms highlights data as a key input into learning and value creation, turning data into a *valuable asset* (Assumption 2). For example, "Facebook, Uber, and Spotify operate technology platforms where their entire value lies in the relationships they create *and* [italics added] the information they hold" (Birkinshaw, 2018, p. 204).

(3) Consumerization (Gabriel, Korczynski, & Rieder, 2015) has blurred the line between consumption and production, turning users into *prosumers* who cocreate value (Assumption 3). For example, content creators on social media platforms such as YouTube simultaneously consume and produce marketing content, effectively cocreating value with brands and other YouTubers.

(4) The fact that a few large platform firms (e.g., Facebook, Google) dominate the information economy by capturing a disproportionate and growing share of the value (Iansiti & Lakhani, 2017) has given rise to concerns about the firms' massive influence. Indeed, platform AI capability alters the behaviors, attitudes, expectations, and emotions of people participating in elections, protests, education, and so forth, affecting the interests of a wide range of stakeholders often in conflicting ways. This suggests that for long-term success, platform owners must *balance diverse stakeholder interests* (Assumption 4).

Drawing on this set of assumptions, we explain our framework in the following sections (Figure 1).

Platform AI Capability

We suggest that the engine driving data network effects is platform AI capability, defined as the ability of a platform to learn from data to continuously improve its products and services for each user (Assumption 1) (Figure 1). The main mechanism through which platform AI capability may enhance perceived user value is by improving prediction (Meinhart, 1966). Prediction describes the ability of a system to draw upon existing data about the past and present to generate information about the future (Churchman, 1961). This information can help forecast future events or provide recommendations for action (Agrawal et al., 2018). For example, a creditworthiness decision made by a lending platform involves predicting the likelihood that someone will pay back a loan, drawing upon existing data on users and past transactions. Another example is the detection of fraudulent credit card transactions, which increasingly relies on machine learning algorithms trained by data scientists and domain experts.

Prediction enabled by machine learning works through what Herbert A. Simon called learning from examples: “A number of systems have been constructed that learn from their own problem-solving efforts, or from the successful problem-solving efforts of others in the form of worked-out examples of problem solutions” (Simon, 1995, p. 110). For example, to develop a reliable fraud detection model as in the example given above, a balanced training dataset with past fraudulent and nonfraudulent examples of credit card transactions must be created and fed into the machine learning algorithm during training.

Under certain circumstances, which include training the machine learning algorithms with adequate datasets, machine-generated predictions can help avoid human cognitive biases in making assessments and forming judgments. For instance, in discussing how to deal with the known overconfidence bias in which a person’s subjective confidence in her judgments is greater than the objective accuracy of those judgments, Kahneman and Tversky (1977) state the following: “The most radical suggestion is to replace such assessments by computation” (p. 4-7).

Effectively, computations enabled by a platform AI capability can result in higher speed and accuracy of prediction (Agrawal et al., 2018). Both types of improvements and their effect on the characteristics of network structure and conduct (Afuah, 2013) have to be taken into account to understand how platform AI capability impacts perceived user value.

Speed of prediction. Users participating in the platform’s network are free agents, empowered by the use of products and services offered by the platform. For example, a Facebook user autonomously decides what to post and when, and an Uber driver decides when, where, and how long to drive and whether to accept or reject ride requests. As a result of the users' autonomy, exchange relationships involving interactions among users are typically bounded in time and affected by a myriad of actions taken at this very moment by other users. For example, Twitter users may retweet messages within seconds, directly influencing other users to engage or disengage in further information exchange on the platform. Such actions may lead to a rapid reconfiguration of the network's structure, which can potentially impede new interactions or manifest in opportunistic behaviors by actors pursuing information asymmetries (Afuah, 2013) and misinformation campaigns (O'Connor & Weatherall, 2019).

A platform AI capability offering a greater speed of prediction helps offset such value-destroying dynamics and foster value-enhancing interactions among users by minimizing the time between when a salient change in the network structure or conduct occurs and when the platform detects this change and generates user action recommendations to influence the network. Indeed, in an ideal scenario, the platform makes instantaneous predictions and anticipates any network dynamics that destroy user value based on the state of the network of users at the exact moment a given transaction is being carried out. For example, Uber tries to prevent fraudulent behavior such as prearranged trips between riders and drivers that limit open competition by letting its algorithms monitor signs of fake trips (e.g., requesting, accepting and completing trips on the same device or with the same payment profile, excessive promotional trips, excessive cancellations) in real time for faster prediction and action recommendations or sanctions to enforce rules more quickly. Similarly, Facebook tries to detect misinformation more quickly to prevent false news from spreading by employing machine learning algorithms that help identify faster what stories might be false or which accounts will more likely post false news before letting human fact-checkers do their work to moderate content and increase the perceived value of the platform.

Proposition 1a: The greater the speed of prediction, the higher the perceived user value is likely to be.

Accuracy of prediction. As illustrated by the examples given above, the learning enabled by platform AI capability not only occurs on the basis of data collected from the network but also influences the network by shaping interactions among users. This influence occurs by wrapping trained prediction models and machine learning features into the products and services offered by the platform, allowing them to function in a smarter and more adaptive way. The resulting agency of the platform exerts a strong influence on key network characteristics, including the perception of trust among network users and transaction feasibility (Afuah, 2013). As an example of the latter, the feasibility of transactions on a platform such as Uber depends not only on the ubiquitous availability and constant connectivity of the Internet and smartphones with installed apps but also on the availability of information generated through prediction (e.g., pushed information about available ride requests on the driver's way home after a platform work shift). As an example of the influence of prediction on trust among interacting users, the information filtering, curation, and ranking performed by algorithms on Facebook has at times generated a greater perception of trust and at other times a weaker perception of trust in the network, depending on the accuracy of the prediction.

As these examples illustrate, a platform AI capability ensuring greater accuracy of prediction helps reduce deviations from what has been forecasted or recommended to what events or outcomes have actually occurred or what users truly want, increasing transaction feasibility and bolstering the perception of trust among network users. For example, when the Uber platform indicates an estimated arrival time of three minutes but it takes the car ten minutes to pick up a customer, the value of the platform to this particular user decreases. Similarly, inaccurate forecasts and misplaced action recommendations due to algorithmic biases (Lambrecht & Tucker, 2019), for example, may fuel malevolent behaviors and lead to a deterioration of trust in the network. Continued difficulties of Facebook algorithms to detect fake news stories offer a good illustration of this latter point (Bucher, 2016).

Proposition 1b: The greater the accuracy of prediction, the higher the perceived user value is likely to be.

Data Stewardship

Data are oftentimes referred to as the oil fueling the information economy (McAfee & Brynjolfsson, 2012; Perrons & Jensen, 2015; Varian, 2014). This suggests that data are a valuable asset (Assumption 2), especially when they are used to nurture platform AI capability and help ensure value creation for each user. When supplied with sufficient quality and quantity of oil, the engine may provide much more value to its users. Similarly, we suggest that the effect of platform AI capability on perceived user value is moderated by data quantity and data quality. To ensure this strengthening effect, a firm must refine and extract value from data by means of data stewardship, defined as the enterprise-wide holistic management of a firm's data assets to help ensure adequate data quantity and quality (Baesens et al., 2016; Cooper, Watson, Wixom, & Goodhue, 2000; Kitchens, Dobolyi, Li, & Abbasi, 2018; Otto, 2011; Ross, Weill, & Robertson, 2006; Wixom & Watson, 2001; Woerner & Wixom, 2015). Data stewardship acts as a mechanism of data network effects by helping fuel the engine, making the platform more valuable to each user through increased speed and accuracy of prediction (Agrawal et al., 2018).

To understand the moderating effect of data quantity and quality on the relationship between platform AI capability and perceived user value, consider the role of data in machine learning, as discussed earlier above. Machine learning algorithms are fed with training data to iteratively adjust predictive models until they produce more accurate and relevant results for the users (Agrawal et al., 2018). Training machine learning algorithms with greater amounts of data leads to better prediction models (Simon, 1995, 1996) from which users will ultimately benefit. However, there are many examples where machine learning algorithms trained on large datasets produce inaccurate prediction results (Khoury & Ioannidis, 2014). For example, IBM's efforts to train machine learning algorithms to diagnose cancer and recommend treatment options, including their probabilities of success, have been greatly complicated by handwritten notes and local acronyms. Accordingly, we suggest that both data quantity and quality need to be considered factors as moderating the impact of platform AI capability on perceived user value.

Data quantity. Increased accuracy and speed of prediction, the main mechanisms through which platform AI capability positively impacts perceived user value, depend on the quantity of data used as an input to train and calibrate machine learning models. In their study of human prediction, Kahneman & Tversky (1977) distinguish between singular information, i.e., data consisting of evidence about a particular case, and distributional information, i.e., base-rate data describing the distribution of outcomes across many cases of the same class. A common reason for inaccurate predictions by a person is the tendency to rely too much on singular information, typically coming from a single case that the person is closely familiar with, and to underweight or ignore distributional information. This is called an internal approach to prediction (Kahneman & Tversky, 1977). To avoid this common bias in prediction, the particular case at hand needs to be compared with the distribution of cases of the same class, thus helping avoid biases in the interpretation of data. This is called an external approach to prediction (Kahneman & Tversky, 1977).

While computers do not suffer from motivational factors or limited cognitive information processing capacities that would make them attached to a particular case (Simon, 1991), the internal approach to prediction may still be present in machine learning if the training dataset is not large enough and does not contain a sufficient range of cases of the same class. This will likely lead to misinterpretation of new cases that the algorithm is confronted with during usage, preventing the fast identification of emerging patterns and accurate predictions (Agrawal et al., 2018). The larger the volume of data about past cases, the greater the ability to build and train machine learning algorithms on a strong distributional dataset that facilitates an external approach to prediction, thereby increasing the accuracy and speed of prediction.

For example, DeepMind's 'AlphaGo' system, which beat the former champion Lee Sedol in the boardgame called Go, a strategy game similar to chess where each player seeks to enclose more territory on the board than the opponent, was trained using vast quantities of examples taken from a large number of games played by the best human Go players, allowing the machine to optimize its prediction and decision-making capabilities to the extent that its speed and accuracy outperformed the best Go player in the world. The limitations of relatively small quantities of data for training machine learning algorithms become apparent in another example. Hedge-fund investors such as AQR Capital Management are increasingly relying on algorithmic trading, using a large variety and volume of data, from credit-card records to satellite images of inventories to flight charters for private jets, to make more accurate predictions and more profitable investment decisions. Yet the overall size of the data relative to the complexity of the events that these hedge fund managers are trying to forecast is still not large enough, highlighting again the importance of data quantity as an important moderator of the relationship between platform AI capability and perceived user value.

Proposition 2a: The higher the quantity of data for the training of machine learning algorithms on the platform, the stronger is the relationship between platform AI capability and perceived user value.

Data quality. Increased accuracy and speed of prediction also depend on the quality of data used as input for training and calibrating machine learning models. Kahneman & Tversky (1977) explain that human predictors typically suffer from an overconfidence bias, whereby their certitude concerning a given estimate tends to be higher than that justified by the available evidence. This happens due to people's tendency to form judgments that are consistent with their preferences and experience as well as due to the adoption of unverified assumptions and due to cognitive anchoring, whereby an individual depends too heavily on an initial piece of information offered when making decisions (Kahneman & Tversky, 1977). While these cognitive limitations, in principle, can be overcome by computation and machine learning, avoiding the overconfidence bias in prediction requires the use of a dataset that is complete, reliable and appropriate for the task at hand (Kahneman & Tversky, 1977). In other words, the data must be of sufficient quality.

Data quality includes truthfulness (the degree of conformity between the recorded value and the actual value), completeness (the extent to which the recorded values exist for all observations), consistency (the degree to which the data have been measured in the same manner across cases), and timeliness (the speed by which data observations are updated in the event of change) (Ballou & Pazer, 1985; Constantiou & Kallinikos, 2015; Markus, 2015; McAfee & Brynjolfsson, 2012; Woerner & Wixom, 2015; Yoo, 2015). The better the quality of data, the greater is the likelihood of reducing or eliminating the prevalent overconfidence bias in prediction (Kahneman & Tversky, 1977), thereby strengthening the impact of platform AI capability on perceived user value.

For example, popular fare aggregators and travel metasearch engines such as Kayak.com offer several alternative routes alongside their prices for users to choose from to reach their desired destination. Making such recommendations, or prescriptions, requires not only generating a prediction for how long a flight sequence might take but also offering an indication of the degree of confidence for the recommendation to buy an airfare ticket for the given destination or to wait until better rates become available on the platform. The more truthful, complete, consistent and timely the dataset that the aggregator platform draws on, achieved, for example, through better integration with the reservation systems of airlines and travel agencies, the faster and better are the predictions and, thus, also the recommendations offered by the platform to each user.

Proposition 2b: The higher the quality of data for the training of machine learning algorithms on the platform, the stronger is the relationship between platform AI capability and perceived user value.

User-Centric Design

The perceived value of a fueled engine is likely to be only as strong as the design of the car in which the engine is installed because the design shapes the experience of the driver. Similarly, platform AI capability trained with adequate quantities and quality of data may help create AI models that provide greater speed and accuracy of prediction from which users may perceive value; the better that these trained AI prediction models are wrapped into well-designed products and services through which users can directly experience the benefits of platform AI capability, the stronger the perceived value of the platform AI capability to each user is likely to be. We argue that to create value for users, firms designing products and services in the era of AI must adapt to consumerization, a process involving the widespread adoption and diffusion of consumer digital technologies by people across society (Gregory, Kaganer, Henfridsson, & Ruch, 2018). By empowering users to cocreate value with their personal data, consumerization blurs the line between consumption and production (Gabriel et al., 2015), effectively turning users into prosumers (Assumption 3). Firms adapt to consumerization by adopting user-centric design, defined as becoming closer to users and better understanding their needs to help increase the performance and effort expectancy of the products and services.

User-centric design involves applying “design to get closer to users and better understand their needs” (Verganti, 2008, p. 436). By better understanding real user needs and designing the platform’s products and services in a way that closely meets their expectations, habits, whims, and desires (Gabriel et al., 2015), user-centric design empowers and engages users to cocreate value by contributing with their feedback and personal data to the ongoing improvement and tuning of AI models and features of the platform. Therefore, user-centric design acts as another key mechanism of data network effects by helping users experience the supplied engine and making the increased speed and accuracy of prediction afforded by platform AI capability more accessible and beneficial to each user. To achieve this outcome, user-centric design must foster user engagement.

One way to conceptualize user engagement on platforms is to consider the intensity with which users interact with the platform’s product and services, ranging from complete avoidance to skilled and committed use (Klein & Sorra, 1996). Platform businesses typically capture user engagement by reporting the number of daily and monthly active users, where “active” corresponds to a certain threshold of committed use with regard to a particular product or service. A high level of committed use across a broad range of a platform’s products and services makes AI-enabled predictions more accurate because it increases the availability of user feedback about the outcome (e.g., user chooses option A) following each instance of forecast or prescription (e.g., user is presented with options A, B and C). Every user interaction with a platform offers an opportunity to test certain features of a product or service and, therefore, to improve the prediction models created by machine learning algorithms to make the user experience more personalized and tailored to the unique identity of each user (Adler & Kwon, 2002). For example, the video streaming service Netflix runs fifty concurrent experiments per user at any given point in time aimed at driving better personalization and continuously developing the feature set of its user applications (Gomez-Uribe & Hunt, 2016). We suggest that both the performance expectancy and effort expectancy of designed products and services need to be considered as factors moderating the impact of platform AI capability on perceived user value by influencing committed use and driving user engagement.

Performance expectancy. To ensure user engagement, the design of the platform's products and services needs to incorporate considerations of performance expectancy. Performance expectancy is defined as the degree to which an individual believes that using the system will help her or him attain gains in job performance (Venkatesh, Morris, Davis, & Davis, 2003). Based on Assumption 3, we view the 'job' in the context of platforms as a series of tasks that the user carries out in a given context by using the platform's products and services (Christensen & Raynor, 2003). The performance will likely be evaluated by users by assessing the extent to which they believe that the adoption of the platform's products and services will help them satisfy their needs and meet their expectations while performing their job. Performance expectancy is the strongest predictor of the user intention to use the system in both voluntary and mandatory settings (Venkatesh et al., 2003) and is therefore a key determinant of committed use (Klein & Sorra, 1996), which is the basis for iterative improvements in predictive models created by machine learning algorithms based on user feedback (Agrawal et al., 2018). Accordingly, performance expectancy is likely to strengthen the impact of platform AI capability on perceived user value.

For example, consider a scenario where a customer needs to travel in a car from point A to point B in the fastest and most convenient way possible under the condition of intense city traffic. Waze, a turn-by-turn navigation app, is enabled by a feature called floating car data (FCD), which determines the traffic speed on the road network based on a collection of local data, speed, direction of travel and time information from mobile phones in vehicles on the road. As this feature is wrapped into the app through user-centric design that induces users to use Waze everyday even though they may know the way to their destination, the app continuously supplies the underlying platform AI capability with new crowdsourced feedback data, helping it improve its predictions on an ongoing basis. As a result, Waze is able to increase the perceived value of the platform for each user through faster and more accurate rerouting based on changing traffic flows.

Proposition 3a: The higher the performance expectancy of the platform's products and services, the stronger is the relationship between platform AI capability and perceived user value.

Effort expectancy. The level of user engagement with a product or service on the platform is also a function of effort expectancy, defined as the degree to which an individual user believes that using the system would be free of effort (Venkatesh et al., 2003). Similar to performance expectancy, effort expectancy also shapes the user intention to adopt the system and, by extension, the level of committed use. The easier it is to use a product or service on the platform, the more likely it is that users will adopt and use it in a committed way, allowing for further data-driven improvement of the underlying AI models and features based on user feedback and creating more value for each user. Thus, user beliefs reflecting higher effort expectancy will likely increase the level of committed use of the platform's products and services, prompting user feedback that is necessary to make the AI-enabled predictions more accurate.

For example, voice assistants such as Apple Siri, Google Now, and Microsoft Cortana found a way to combine complex machine learning technology—deep neutral networks, hybrid emotional inference models as well as natural language processing and generation—with highly accessible user interface designs that rely on voice interaction as a more natural and intuitive way for humans to interact with the machines and use products and services on the respective platforms of each voice assistant service (e.g., Apple iOS). As a result, the perceived effort expectancy of these voice assistant services is very high, contributing to their widespread adoption and engaged use, which in turn helps continuously improve predictions and behavior on the basis of user feedback data that increases the perceived value of the platform.

Proposition 3b: The higher the effort expectancy of the platform's products and services, the stronger is the relationship between platform AI capability and perceived user value.

Platform Legitimation

A car may be nicely designed and powered by a good engine supplied with sufficient quantities of high-quality oil, but people will still only want to use the car if they also consider it safe and secure and the perceived risk of an accident as low. Drawing on this analogy, platform owners must balance diverse stakeholder interests (Assumption 4) to mitigate the perceived risks related to data privacy and security (Cavoukian & Chanliau, 2013; Kroener & Wright, 2014) as well as the interpretability and explainability of AI (Coglianese & Lehr, 2019). Building upon this assumption, we introduce the third key mechanism of data network effects. We argue that actions, including the responsible use of data and ensuring the explainability of AI features, must be considered strategic, as they may play an important role in strengthening the relationship between platform AI capability and perceived user value by avoiding accidents such as data security breaches, data privacy violations, and unintended consequences of unexplainable machine behavior. To capture this category of actions geared toward balancing diverse stakeholder interests and mitigating the perceived risks of the use of big data and AI in platform contexts, we introduce the concept of platform legitimation, defined as actions that the platform owner takes to ensure positive legitimacy evaluations of the platform by key stakeholder audiences. In what follows, we explain the moderating role of platform legitimation in our model.

Legitimacy, defined as "a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, beliefs, and definitions" (Suchman, 1995, p. 574), acts as a key determinant of a social entity's ability to acquire resources from the environment (Garud, Schildt, & Lant, 2014; Zimmerman & Zeitz, 2002). The crucial resources in the case of platforms in today's era of AI include the personal data, financial means, and technological capabilities needed to set up machine learning algorithms, train models, and develop new platform features. Accordingly, regulators overseeing the use of personal data, platform investors, and technology partners all represent key stakeholder groups whose legitimacy judgments must be considered in understanding the functioning of data network effects and ultimately the perceived value of the platform by users.

Satisfying the needs and interests of these key stakeholder groups is important for platform owners because they provide critical resources (e.g., sustained access to personal data protected by appropriate laws and rules) upon which the continued development and use of their platform AI capability depends. The key characteristics of the platform that attract legitimation scrutiny from resource-granting stakeholders and that therefore must be proactively addressed as part of platform legitimation include (1) how the platform is designed and governed to collect, store and use personal data and (2) how the platform is designed and governed to apply machine learning transparently and make predictions explainable.

Personal data use. A critical aspect of platform legitimation concerns the extent to which the platform's approach to collecting, storing, and sharing personal user data is adjudged by the stakeholder audiences to be "the right thing to do" (Suchman, 1995, p. 579). This assessment goes beyond self-interested calculations concerning the utility of platform transactions for an individual user and involves considerations of moral desirability entertained by a wide range of stakeholders across society (Bitektine, 2011). To this end, the platform firm must demonstrate that its policies and procedures for data collection and use, typically communicated through user privacy policies (Bélanger, 2011; Hong, 2013; Pavlou, 2011; Smith, 2011) and information security compliance documents (Anderson & Moore, 2006; Barlow, Warkentin, Ormond, & Dennis, 2018), meet morally desirable principles, such as privacy-by-design and security-by-design (Cavoukian & Chanliau, 2013; Kroener & Wright, 2014). These by-design principles call for data privacy and security to be taken into account throughout the entire engineering and development process and for them to be reflected in the design of the platform or specific products and services on the platform. The declared policies, procedures, and design choices can then be compared by the resource-granting stakeholder audiences with the actual platform outcomes to uncover inconsistencies or malfeasance in how the management applies the norms in practice. In case the platform design and outcomes are deemed incoherent, the regulators, investors and partners may choose to withhold legitimacy and, by extension, resources, forcing the platform to alter or altogether eliminate certain AI features or models.

For example, Uber's expansion into Europe resulted in a backlash against the company's alleged noncompliance with the regional personal data protection regulations as well as, more broadly, against Uber's practices in using city transportation data. Facebook, too, has repeatedly attracted legitimation scrutiny by key stakeholder audiences, including regulators, investors, and partners, over its repeated failures to ensure the privacy and security of user data (e.g., the Cambridge Analytica scandal, where the company was able to harvest personally identifiable information from the Facebook platform through an app that exploited the Facebook Graph API), pointing to limitations in the design and governance of its platform. As a sign of platform legitimation and effort to secure support from key stakeholder groups to sustain its scalable business model around the use of platform AI capability to create value for billions of users and attract advertisers, Facebook has started to endorse data privacy protection rules and to work with regulators to secure positive legitimacy evaluations in the future.

Proposition 4a: The higher the moral desirability of the use of personal data by the platform, the stronger is the relationship between platform AI capability and perceived user value.

Prediction explainability. Another critical aspect of platform legitimation concerns explainability, i.e., interpretability of functioning and coherence in understanding, of the predictions made by AI models and features on the platform. AI-made predictions not only influence core market-related processes on the platform, including how the platform matches different user groups, but also have a profound effect on the behavior and emotions of users. As resource-granting stakeholders seek an understanding of how and why people are being influenced and are affected by these AI-made predictions and the resulting machine behavior or decision making, the stakeholders make an assessment as to whether they are meaningful in the context of the prevalent beliefs, logics, and categories (Suchman, 1995). Considering the "black box" nature of many AI models, which makes it difficult, if not impossible, for humans to understand exactly how machine learning algorithms make predictions and arrive at certain decisions, recommendations, or behaviors (Coglianese & Lehr, 2019), making such predictions explainable is extremely difficult in some cases (Mayenberger, 2019; Preece, 2018). However, only if the explainability of AI-made predictions is achieved can stakeholders assess the meaningfulness of these predictions and renew their trust and commitment to grant the critical resources that help ensure a strong relationship between platform AI capability and perceived user value (Rossi, 2018).

For example, it is becoming increasingly common that banks and lenders use machine learning algorithms to predict credit risk and make creditworthiness assessments. The resulting loan decisions may have a strong impact on the lives of consumers, but disappointed users typically lack an explanation for being denied credit. To increase the perceived user value of AI-enabled loan decision making, credit-granting institutions can educate their customers. For instance, Bank of America offers all customers their FICO (Fair Isaac Corporation) score and explains the important components of the score that are calculated by the algorithms. Thus, fostering explainability of predictions made by machine learning algorithms on the platform is likely to strengthen the relationship between platform AI capability and perceived user value.

Proposition 4b: The higher the explainability of predictions made by machine learning algorithms on the platform, the stronger is the relationship between platform AI capability and perceived user value.

DISCUSSION

Our research contributes to the literature on network effects (Afuah, 2013; Cennamo & Santalo, 2013; Fuentelsaz, Maicas, & Polo, 2012; Gallaugher & Wang, 2002; Liu, Gal-Or, Kemerer, & Smith, 2011; Parker, Van Alstyne, & Jiang, 2017; Shankar & Bayus, 2003;

Sheremata, 2004; Singh, Tan, & Mookerjee, 2011; Suarez, 2005) by explaining the role of AI and data network effects for creating user value, especially in the context of multi-sided platforms (Hagiu & Wright, 2015; Mcintyre et al., 2020). Data network effects exhibit a positive direct relationship between the AI capability of a platform and the value perceived in the platform by its users—a relationship that is moderated by platform legitimation, data stewardship and user-centric design. This highlights new platform externalities, where a user's utility of a platform is a function of the scale of data-driven learning and improvements realized with AI, complementing user value rooted in network externalities deriving from the scale of the network. Integrating our model of data network effects with the extant network effects literature, we argue that the utility that a user derives from a platform is increasingly both a function of the scale of the network and data-driven learning and improvements realized with AI. This highlights the need to examine interactions between network effects and data network effects.

Our explanation of user value creation in the era of AI offers novel set of insights. First, the research describes data network effects as a new category of network effects focused on the impact of data-driven learning and improvements, enabled by platform AI capability, on perceived user value. Under certain conditions, data network effects play an influential role for the value that users perceive in a platform, product, or service. We surmise that data network effects largely influence perceived user value in the context of platforms facilitating the production and exchange of information or experience goods (Shapiro & Varian, 1999; Varian, 2014). In such contexts, the user experience is heavily shaped by the scale of learning from data collected on users. For example, Google Search is an online service powered by a platform AI capability enjoying a high popularity among users due to its capability to continuously improve the underlying algorithms and experience of each user as it learns about users and their search queries. Similarly, Netflix leverages data network effects as it collects and analyzes data about how its platform is used and then draws on the learning outcomes to continuously improve its content and user interface to increase the perceived value of the streaming services offering through its platform. Notably, data network effects are even more significant when learning capabilities are an important determinant of platform, product, or service quality (Mcintyre & Srinivasan, 2017; Zhu & Iansiti, 2012). For example, the user value of a Tesla car's Autopilot functionality is influenced by the firm's ability to use AI and learn from the data collected from sensors, cameras, and radar units in cars to continuously improve the self-driving algorithms and Autopilot functionality.

Second, the value that users perceive in a platform, product, or service may depend on combinations of data network effects and direct network effects. Direct network effects describe the value that users derive from a network, which comes from users being able to interact directly with each other (Katz & Shapiro, 1985; Rochet & Tirole, 2003). In view of the "computer in the middle of every transaction" (Varian, 2014, p. 1), these direct exchanges among users are increasingly mediated by interactive processes of learning from data collected on each user participating in the exchange relationship, highlighting the combined contribution of data network effects and direct network efforts to user value. For example, Facebook has historically benefitted from strong direct network effects, whereby the value that users derive from the network primarily stems from the opportunities of users to interact directly with each other. More recently, however, the "self-reinforcing process whereby growth begets growth" (Boudreau & Jeppesen, 2015, p. 1774) seems to have slowed down, and Facebook has struggled to sustain high-quality interactions among users in increasingly crowded social networks. To deal with this challenge and sustain the perceived user value of the platform, Facebook has activated and started leveraging data network effects on top of direct network effects by collecting and learning from the vast amounts of personal data from its large "N" of users on the network (Farrell & Saloner, 1986; Gandal, 1994; Katz & Shapiro, 1985). By applying machine learning techniques and rolling out AI models and features, Facebook has tried to influence the network in a desirable direction to increase perceived user value. The increase in perceived user value stemming from data network effects feeds back into direct network effects, as it increases the number of daily active users, offering more opportunities for users to interact directly with each other.

Third, the value that users perceive in a platform or one of its products or services may depend on combinations of data network effects and indirect network effects. Indirect network effects focus on the phenomenon that the more people that use a product, the greater is the variety and availability of the complements of the product, thereby increasing perceived user value (Boudreau, 2012; Church et al., 2008; Clements & Ohashi, 2005). This phenomenon of the demand for a product and the supply of complements for that product affecting each other (Stremersch, Tellis, Franses, & Binken, 2007) may be influenced by data network effects if the developers of complements can use the platform AI capability to learn from data collected on users of the product to improve the quality of their complements. For example, Apple rolled out an AI model framework for iOS developers (called Core ML), bringing machine learning to smartphone apps in its mobile ecosystem. While each user of Apple's mobile ecosystem benefitted before from indirect network effects that resulted in a greater diversity and number of complements of iPhones and other iOS devices, these indirect network effects are now strengthened by data network effects, as developers are provided by Apple with a platform AI capability that helps them improve their apps by performing fast and accurate predictions, potentially increasing the perceived value of the complements and overall mobile ecosystem for each user. Examples of such improvements include real-time image recognition, face detection, text prediction, and speaker identification. As an increasing number of these kinds of features enabled by Apple's platform AI capability are incorporated into complements of iPhones and other iOS devices, perceived user value is increasingly becoming a function of the combination of data network effects and indirect network effects.

Finally, our research indicates the significance of extending the scope of network effects research beyond the economics view of platforms (cf. Gawer, 2014). Data network effects relate to the technical architecture of the platform, indicating that network effects theory needs to go beyond viewing platforms as mere markets (cf. Mcintyre et al., 2020) to effectively study how network effects interact with data network effects. The research reported in this paper indicates some of the factors that need to be incorporated into our understanding of how network effects are empowered by AI technologies (cf. Raisch & Krakowski, 2020).

For future research, we suggest empirically examining data network effects in the context of direct and/or indirect network effects. In doing so, it makes sense to distinguish between positive and negative data network effects, similar to the common distinction between positive and negative direct or indirect network effects (Parker, Van Alstyne, & Choudary, 2016). The relevance of making this distinction is highlighted by the nature of machine learning. Machine learning algorithms at work in data network effects learn on the basis of data, yielding unique models of prediction and decision making. This phenomenon is also referred to as self-programming, in contrast with knowledge-based systems that are explicitly programmed (Meinhart, 1966; Samuel, 1959). Such self-programming also has disadvantages, including the possibility of algorithmic biases (Lambrecht & Tucker, 2019). We therefore suggest distinguishing between positive and negative data network effects (cf. Parker et al., 2016). Negative data network effects, where the perceived value of the platform for users decreases, may particularly be activated in the absence of high data quality and quantity as well as during breaches of data privacy and security (see model in Figure 1). As an example of the former scenario, consider Microsoft's AI-powered chatbot Tay, a Twitter bot that was supposed to learn to engage people through casual and playful conversations on social media. Tay rapidly picked up racist and highly abusive language from Twitter users, causing a rapid deterioration of perceived user value. As this example illustrates, embedding platform AI capabilities in exchange relationships and user networks on multi-sided platforms poses considerable risks (Russell, Hauert, Altman, & Veloso, 2015), highlighting the need to consider both the intended and unintended consequences of data network effects in future research.

Future work could also explore the impact of AI and data network effects on the value of the platform to the platform owner. By expanding the focus of explanation from perceived user value (our model) to value creation and capture, future work could effectively explore the linkages between data network effects and competitive advantage. For example, how and why do 'superior' data-driven AI processes in a firm erode the traditional isolating mechanisms that incumbent leaders might have built into an industry?[1] One of the most significant isolating mechanisms discussed in prior strategic management literature is the firm's idiosyncratic capacity to learn and diversify at the same time (Kor & Mahoney, 2004). In platform AI settings, learning processes are data-driven, while diversification oftentimes involves diversifying the platform in a way that stimulates new related application areas (Cennamo & Santaló, 2018; Ghazawneh & Henfridsson, 2013). The platform sponsor's idiosyncratic capacity to learn from data is enabled by complementarities between the platform AI capability and various management, governance and design capabilities by the platform organization (see our model); these complementarities, when

[1] We thank an anonymous reviewer for providing this suggestion.

paired with the relatedness of platform-based products and services, may lead to a generative diversification of the platform that becomes increasingly difficult to imitate over time. The extent to which this unique isolating mechanism in platform AI settings erodes the traditional isolating mechanisms in established industries is likely to depend on the relative role of learning from data and information compared to other critical success factors of competition.

Finally, future research could also explore the interaction between artificial and collective intelligence. This may involve studying the usefulness of AI in evaluating solutions from crowds, which is particularly relevant when there are many solutions and no one knows what the best solution should be (Afuah & Tucci, 2012; Piezunka & Dahlander, 2015). Furthermore, understanding how to best organize work by iteratively leveraging artificial and collective intelligence, as well as combining them, can help in expediting search processes, completing modular and complex work, or helping to identify optimal solutions (Afuah & Tucci, 2012; Baldwin & Clark, 2000; Yu & Nickerson, 2011). On the one hand, AI can augment distributed problem solving and production models (Kyriakou, Nickerson, & Sabnis, 2017), while on the other hand, crowds can support AI systems by providing skills that these systems currently lack (Kittur et al., 2013; Von Ahn & Dabbish, 2004).

IMPLICATIONS FOR MANAGERS

Managers are well aware that data are the new oil fueling the information economy and that data should be treated as a strategic asset (McAfee & Brynjolfsson, 2012; Perrons & Jensen, 2015; Varian, 2014). Reaping the strategic benefits from data assets requires, in addition to the development or acquisition of a superior platform AI capability, careful attention to three key mechanisms of data network effects: (1) data stewardship, (2) user-centric design, and (3) platform legitimation. In terms of (1), this means ensuring that machine learning algorithms on the platform are fed with appropriate quantities and quality of data and that they employ an enterprise-wide approach to the holistic governance of the firm's data assets. In terms of (2), this means embracing consumerization during development to create user-centric designs of products and services on the platform that increase performance expectancy (e.g., greater personalization) and effort expectancy (e.g., greater ease-of-use). In terms of (3), this means using personal data collected from users responsibly by implementing principles of privacy-by-design and security-by-design and by ensuring the explainability of predictions generated by AI on the platform. When all these mechanisms are successfully activated, users will likely perceive sustainable value in the platform AI capability, which may then become a source of competitive advantage.

REFERENCES

Adler, P. S., & Kwon, S.-W. 2002. Social Capital: Prospects for a New Concept. ***Academy of Management Review***, 27(1): 17-40.

Afuah, A. 2013. Are network effects really all about size? The role of structure and conduct. ***Strategic Management Journal***, 34(3): 257-273.

Afuah, A., & Tucci, C. L. 2012. Crowdsourcing As a Solution to Distant Search. ***Academy of Management Review***, 37(3): 355-375.

Agrawal, A., Gans, J., & Goldfarb, A. 2018. ***Prediction Machines: The Simple Economics of Artificial Intelligence***. Boston, MA: Harvard Business Review.

Ahsen, M. E., Ayvaci, M. U. S., & Raghunathan, S. 2019. When Algorithmic Predictions Use Human-Generated Data: A Bias-Aware Classification Algorithm for Breast Cancer Diagnosis. ***Information Systems Research***, 30(1): 97-116.

Anderson, R., & Moore, T. 2006. The Economics of Information Security. ***Science***, 314(5799): 610-613.

Arthur, W. B. 2009. ***The Nature of Technology: What it is and How it Evolves***: Free Press.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. 2016. Transformational Issues of Big Data and Analytics in Networked Business. ***MIS Quarterly***, 40(4): 807-818.

Baldwin, C. Y., & Clark, K. B. 2000. ***Design rules: The power of modularity***: MIT press.

Ballou, D. P., & Pazer, H. L. 1985. Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems. ***Management Science***, 31(2): 150-162.

Barlow, J. B., Warkentin, M., Ormond, D., & Dennis, A. R. 2018. Don't Even Think About It! The Effects of Antineutralization, Informational, and Normative Communication on Information Security Compliance. ***Journal of the Association for Information Systems***, 19: 689-715.

Bélanger, F. C., Robert E. 2011. Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. ***MIS Quarterly***, 35(4): 1017-A1036.

Birkinshaw, J. 2018. How is technological change affecting the nature of the corporation? ***Journal of the British Academy***, 6(s1): 185-214.

Bitektine, A. 2011. Toward a Theory of Social Judgments of Organizations: The Case of Legitimacy, Reputation, and Status. ***The Academy of Management Review***, 36(1): 151-179.

Boudreau, K. J. 2012. Let a Thousand Flowers Bloom? An Early Look at Large Numbers of Software App Developers and Patterns of Innovation. ***Organization Science***, 23(5): 1409-1427.

Boudreau, K. J., & Jeppesen, L. B. 2015. Unpaid Crowd Complementors: The Platform Network Effect Mirage. ***Strategic Management Journal***, 36(12): 1761-1777.

Bucher, T. 2016. The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. ***Information, Communication & Society***, 20(1): 30-44.

Cavoukian, A., & Chanliau, M. 2013. ***Privacy and security by design: A convergence of paradigms***. Toronto, Canada: Information and Privacy Commissioner.

Cennamo, C., & Santalo, J. 2013. Platform Competition: Strategic Trade-Offs in Platform Markets. ***Strategic Management Journal***, 34(11): 1331-1350.

Cennamo, C., & Santaló, J. 2018. Generativity Tension and Value Creation in Platform-based Technology Ecosystems. ***Organization Science***, Forthcoming.

Chen, D. L., & Horton, J. J. 2016. Research Note—Are Online Labor Markets Spot Markets for Tasks? A Field Experiment on the Behavioral Response to Wage Cuts. ***Information Systems Research***, 27(2): 403-423.

Christensen, C. M., & Raynor, M. E. 2003. ***The Innovator's Solution: Creating and Sustaining Successful Growth***. Cambridge, MA: Harvard Business School Press.

Church, J., & Gandal, N. 1992. Network Effects, Software Provision, and Standardization. ***The Journal of Industrial Economics***, 40(1): 85-103.

Church, J., Gandal, N., Krause, D., & Canada, B. 2008. Indirect Network Effects and Adoption Externalities. ***Review of Network Economics***, 7(3).

Churchman, C. W. 1961. ***Predicteion and optimal decision***. Englewood Cliffs, N.J.: Prentice-Hall.

Clements, M., & Ohashi, H. 2005. Indirect Network Effects and the Product Cycle: Video Games in the U.S., 1994-2002. ***The Journal of Industrial Economics***, 53(4): 515-542.

Clements, M. T. 2004. Direct and indirect network effects: are they equivalent? ***International Journal of Industrial Organization***, 22(5): 633-645.

Coglianese, C., & Lehr, D. 2019. Transparency and Algorithmic Governance. ***Administrative Law Review***, 71(1): 1-56.

Constantiou, I. D., & Kallinikos, J. 2015. New games, new rules: big data and the changing context of strategy. ***Journal of Information Technology***, 30(1): 44-57.

Cooper, B. L., Watson, H. J., Wixom, B. H., & Goodhue, D. L. 2000. Data Warehousing Supports Corporate Strategy at First American Corporation. ***MIS Quarterly***, 24(4): 547-567.

Eisenmann, T., Parker, G., & Van Alstyne, M. 2011. Platform Envelopment. ***Strategic Management Journal***, 32(12): 1270-1285.

Farrell, J., & Saloner, G. 1985. Standardization, compatibility, and innovation. ***RAND Journal of Economics***, 16: 70-83.

Farrell, J., & Saloner, G. 1986. Installed base and compatibility: innovation, product pre-announcement, and predation. ***American Economic Review***, 76(5): 940-955.

Finlay, S. 2017. ***Artificial Intelligence and Machine Learning for Business***. Lancashire: Relativistic Books.

Fuentelsaz, L., Maicas, J. P., & Polo, Y. 2012. Switching Costs, Network Effects, and Competition in the European Mobile Telecommunications Industry. ***Information Systems Research***, 23(1): 93-108.

Gabriel, Y., Korczynski, M., & Rieder, K. 2015. Organizations and their Consumers: Bridging Work and Consumption. ***Organization***, 22(5): 629-643.

Gallaugher, J. M., & Wang, Y.-M. 2002. Understanding Network Effects in Software Markets: Evidence from Web Server Pricing. ***MIS Quarterly***, 26(4): 303-327.

Gandal, N. 1994. Hedonic Price Indexes for Spreadsheets and an Empirical Test for Network Externalities. ***The RAND Journal of Economics***, 25(1): 160-170.

Garud, R., & Karnøe, P. 2003. Bricolage versus breakthrough: distributed and embedded agency in technology entrepreneurship. ***Research Policy***, 32(2): 277-300.

Garud, R., Schildt, H. A., & Lant, T. K. 2014. Entrepreneurial Storytelling, Future Expectations, and the Paradox of Legitimacy. ***Organization Science***, 25(5): 1479–1492.

Gawer, A. 2009. ***Platforms, Markets, and Innovation***. Cheltenham, Gloucestershire, UK: Edward Elgar.

Gawer, A. 2014. Bridging differing perspectives on technological platforms: Toward an integrative framework. ***Research Policy***, 43(7): 1239-1249.

Ghazawneh, A., & Henfridsson, O. 2013. Balancing Platform Control and External Contribution in Third-Party Development: The Boundary Resources Model. ***Information Systems Journal***, 23(2): 173-192.

Gomez-Uribe, C. A., & Hunt, N. 2016. The Netflix Recommender System: Algorithms, Business Value, and Innovation. ***ACM Transactions on Management Information Systems (TMIS)***, 6(4): Article No. 13.

Gregory, R. W., Kaganer, E., Henfridsson, O., & Ruch, T. J. 2018. IT Consumerization and the Transformation of IT Governance. ***MIS Quarterly***, 42(4): 1225-1253.

Hagiu, A., & Wright, J. 2015. Multi-sided platforms. ***International Journal of Industrial Organization***, 43: 162-174.
Himan, N. B. 2002. Platform Externalities and the Antitrust Case Against Microsoft. ***The Antitrust Bulletin***.
Hong, W. L. T., James Y. 2013. Internet Privacy Concerns: An Integrated Conceptualization and Four Empirical Studies. ***MIS Quarterly***, 37(1): 275-298.
Iansiti, M., & Lakhani, K. R. 2017. Managing Our Hub Economy. ***Harvard Business Review***, 95(5): 84-92.
Kahneman, D., & Tversky, A. 1977. Intuitive Prediction: Biases and Corrective Procedures: Decisions and Designs Inc.
Katz, M. L., & Shapiro, C. 1985. Network Externalities, Competition, and Compatibility. ***The American Economic Review***, 75(3): 424-440.
Katz, M. L., & Shapiro, C. 1986. Technology Adoption in the Presence of Network Externalities. ***Journal of Political Economy***, 94(4): 822-841.
Katz, M. L., & Shapiro, C. 1992. Product Introduction with Network Externalities. ***The Journal of Industrial Economics***, 40(1): 55-83.
Khoury, M. J., & Ioannidis, J. P. A. 2014. Big Data Meets Public Health. ***Science***, 346(6213): 1054-1055.
Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. 2018. Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data. ***Journal of Management Information Systems***, 35(2): 540-574.
Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., & Horton, J. 2013. ***The future of crowd work***.
Klein, K. J., & Sorra, J. S. 1996. The Challenge of Innovation Implementation. ***The Academy of Management Review***, 21(4): 1055-1080.
Kor, Y. Y., & Mahoney, J. T. 2004. Edith Penrose's (1959) Contributions to the Resource-based View of Strategic Management. ***Journal of Management Studies***, 41(1): 183-191.
Kroener, I., & Wright, D. 2014. A Strategy for Operationalizing Privacy by Design. ***The Information Society***, 30(5): 355-365.
Kyriakou, H., Nickerson, J. V., & Sabnis, G. 2017. Knowledge Reuse for Customization: Metamodels in an Open Design Community for 3D Printing. ***MIS Quarterly***, 41(1): 315-322.
Lambrecht, A., & Tucker, C. 2019. Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. ***Management Science***, 65(7): 2966-2981.
Liu, C. Z., Gal-Or, E., Kemerer, C., & Smith, M. D. 2011. Compatibility and Proprietary Standards: The Impact of Conversion Technologies in IT Markets with Network Effects. ***Information Systems Research***, 22(1): 188-207.
Majumdar, S. K., & Venkataraman, S. 1998. Network Effects and the Adoption of New Technology: Evidence From the U.S. Telecommunications Industry. ***Strategic Management Journal***, 19(11): 1045-1062.
Markus, M. L. 2015. New games, new rules, new scoreboards: the potential consequences of big data. ***Journal of Information Technology***, 30(1): 58-59.
Mayenberger, D. 2019. How to overcome modelling and model risk management challenges with artificial intelligence and machine learning. ***Journal of Risk Management in Financial Institutions***, 12(3): 241-255.
McAfee, A., & Brynjolfsson, E. 2012. Big Data: The Management Revolution. ***Harvard Business Review***, 90(10): 60-68.
McCorduck, P. 2004. ***Machines Who Think: A Personal Inquiry into the History and Prospect of Artificial Intelligence***. Natick, Massachusetts: A K Peters/CRC Press; 2 edition.
Mcintyre, D. P., & Srinivasan, A. 2017. Networks, Platforms, and Strategy: Emerging Views and Next Steps. ***Strategic Management Journal***, 38(1): 141-160.
Mcintyre, D. P., Srinivasan, A., Afuah, A., Gawer, A., & Krestschmer, T. 2020. Multi-sided Platforms as New Organizational Forms: A Dynamic Perspective on Strategy, Scope and Business Models. ***Academy of Management Perspectives***.
Meinhart, W. A. 1966. Artificial Intelligence, Computer Simulation of Human Cognitive and Social Processes, and Management Thought. ***The Academy of Management Journal***, 9(4): 294-307.
Meyer, G., Adomavicius, G., Johnson, P. E., Elidrisi, M., Rush, W. A., Sperl-Hillen, J. M., & O'Connor, P. J. 2014. A Machine Learning Approach to Improving Dynamic Decision Making. ***Information Systems Research***, 25(2): 239-263.
O'Connor, C., & Weatherall, J. O. 2019. ***The Misinformation Age: How False Beliefs Spread***. New Haven, Connecticut: Yale University Press.
Otto, B. 2011. Data Governance. ***Business & Information Systems Engineering***, 3(4): 241-244.

Parker, G., & Van Alstyne, M. 2018. Innovation, Openness, and Platform Control. ***Management Science***, 64(7): 3015-3032.

Parker, G., Van Alstyne, M., & Jiang, X. 2017. Platform Ecosystems: How Developers Invert the Firm. ***MIS Quarterly***, 41(1): 255-266.

Parker, G. G., & Van Alstyne, M. 2005. Two-Sided Network Effects: A Theory of Information Product Design. ***Management Science***, 51(10): 1494-1504.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. 2016. ***Platform revolution: How networked markets are transforming the economy--and how to make them work for you*** (Kindle ed.): WW Norton & Company.

Pavlou, P. 2011. State of the Information Privacy Literature: Where are We Now and Where Should We Go? ***MIS Quarterly***, 35(4): 977-988.

Perrons, R. K., & Jensen, J. W. 2015. Data as an asset: What the oil and gas sector can learn from other industries about "Big Data". ***Energy Policy***, 81: 117-121.

Piezunka, H., & Dahlander, L. 2015. Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. ***Academy of Management Journal***, 58(3): 856-880.

Preece, A. 2018. Asking 'Why' in AI: Explainability of intelligent systems - perspectives and challenges. ***Intelligent Systems in Accounting, Finance and Management***, 25(2): 63-72.

Priem, R. L., Butler, J. E., & Li, S. 2013. Toward Reimagining Strategy Research: Retrospection and Prospection on the 2011 AMR Decade Award Article. ***Academy of Management Review***, 38(4): 471-489.

Raisch, S., & Krakowski, S. 2020. Artificial Intelligence and Management: The Automation-Augmentation Paradox. ***Academy of Management Review***.

Rochet, J.-C., & Tirole, J. 2003. Platform Competition in Two-Sided Markets. ***Journal of the European Economic Association***, 1(4): 990-1029.

Rochet, J.-C., & Tirole, J. 2006. Two-Sided Markets: A Progress Report. ***The RAND Journal of Economics***, 37(3): 645-667.

Rosenblat, A. 2018. ***Uberland: How Algorithms Are Rewriting the Rules of Work***. Berkeley, California: University of California Press.

Ross, J. W., Weill, P., & Robertson, D. 2006. ***Enterprise Architecture as Strategy: Creating a Foundation for Business Execution***. Boston: Harvard Business Press.

Rossi, F. 2018. Building Trust in Artificial Intelligence. ***Journal of International Affairs***, 72(1): 127-133.

Russell, S., Hauert, S., Altman, R., & Veloso, M. 2015. Ethics of artificial intelligence. ***Nature***, 521(7553): 415-416.

Samuel, A. L. 1959. Some Studies in Machine Learning Using the Game of Checkers. ***IBM Journal of Research and Development***, 3(3): 210-229.

Schilling, M. A. 2000. Toward a General Modular Systems Theory and Its Application to Interfirm Product Modularity. ***The Academy of Management Review***, 25(2): 312-334.

Schilling, M. A. 2002. Technology Success and Failure in Winner-Take-All Markets: The Impact of Learning Orientation, Timing, and Network Externalities. ***The Academy of Management Journal***, 45(2): 387-398.

Shankar, V., & Bayus, B. L. 2003. Network Effects and Competition: An Empirical Analysis of the Home Video Game Industry. ***Strategic Management Journal***, 24(4): 375-384.

Shapiro, C., & Varian, H. R. 1999. ***Information Rules***. Cambridge, MA: Harvard Business School Press.

Sheremata, W. A. 2004. Competing Through Innovation in Network Markets: Strategies for Challengers. ***Academy of Management Review***, 29(3): 359-377.

Simon, H. A. 1991. Bounded Rationality and Organizational Learning. ***Organization Science***, 2(1): 125-134.

Simon, H. A. 1995. Artificial Intelligence: An Empirical Science. ***Artificial Intelligence***, 77: 95-127.

Simon, H. A. 1996. ***The Sciences of the Artificial*** (3rd ed.). Cambridge, MA and London, England: The MIT Press.

Singh, P. V., Tan, Y., & Mookerjee, V. 2011. Network Effects: The Influence of Structural Capital on Open Source Project Success. ***MIS Quarterly***, 35(4): 813-A817.

Smith, H. J. D., Tamara; Xu, Heng. 2011. Information Privacy Research: An Interdisciplinary Review. ***MIS Quarterly***, 35(4): 980-A927.

Stremersch, S., Tellis, G. J., Franses, P. H., & Binken, J. L. G. 2007. Indirect Network Effects in New Product Growth. ***Journal of Marketing***, 71(3): 52-74.

Stucke, M. E., & Ezrachi, A. 2016. When Competition Fails to Optimize Quality: A Look at Search Engines. ***Yale Journal of Law and Technology***, 18: 70-110.

Stucke, M. E., & Ezrachi, A. 2018. Alexa et al., What Are You Doing with My Data? ***New Economic Analysis of Law***, 5(1).

Suarez, F. F. 2005. Network Effects Revisited: The Role of Strong Ties in Technology Selection. ***Academy of Management Journal***, 48(4): 710-720.

Suchman, M. C. 1995. Managing legitimacy: Strategic and institutional approaches. ***Academy of Management Review***, 20(3): 571–610.

Tucker, C. 2019. Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs, Essential Facility. ***Review of Industrial Organization***, 54(4): 683-694.

Van Alstyne, M. W., Parker, G. G., & Choudary, S. P. 2016. Pipelines, Platforms, and the New Rules of Strategy. ***Harvard Business Review***, April.

Varian, H. R. 2014. Beyond Big Data. ***Business Economics***, 49(1): 27-31.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. 2003. User Acceptance of Information Technology: Toward a Unified View. ***MIS Quarterly***, 27(3): 425-478.

Verganti, R. 2008. Design, Meanings, and Radical Innovation: A Metamodel and a Research Agenda. ***Journal of Product Innovation Management***, 25(5): 436-456.

Von Ahn, L., & Dabbish, L. 2004. ***Labeling images with a computer game***. Paper presented at the Conference on Human Factors in Computing Systems.

Wixom, B. H., & Watson, H. J. 2001. An Empirical Investigation of the Factors Affecting Data Warehousing Success. ***MIS Quarterly***, 25(1): 17-41.

Woerner, S. L., & Wixom, B. H. 2015. Big Data: Extending the Business Strategy Toolbox. ***Journal of Information Technology***, 30(1): 60-62.

Yoo, Y. 2015. It is not about size: a further thought on big data. ***Journal of Information Technology***, 30(1): 63-65.

Yoo, Y., Henfridsson, O., & Lyytinen, K. 2010. Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. ***Information Systems Research***, 21(4): 724-735.

Yu, L., & Nickerson, J. V. 2011. ***Cooks or cobblers?: crowd creativity through combination***.

Zhu, F., & Iansiti, M. 2012. Entry Into Platform-Based Markets. ***Strategic Management Journal***, 33(1): 88-106.

Zhu, F., & Liu, Q. 2018. Competing with complementors: An empirical look at Amazon.com. ***Strategic Management Journal***, 39(10): 2618-2642.

Zimmerman, M. A., & Zeitz, G. J. 2002. Beyond survival: Achieving new venture growth by building legitimacy. ***Academy of Management Review***, 27(3): 414–431.

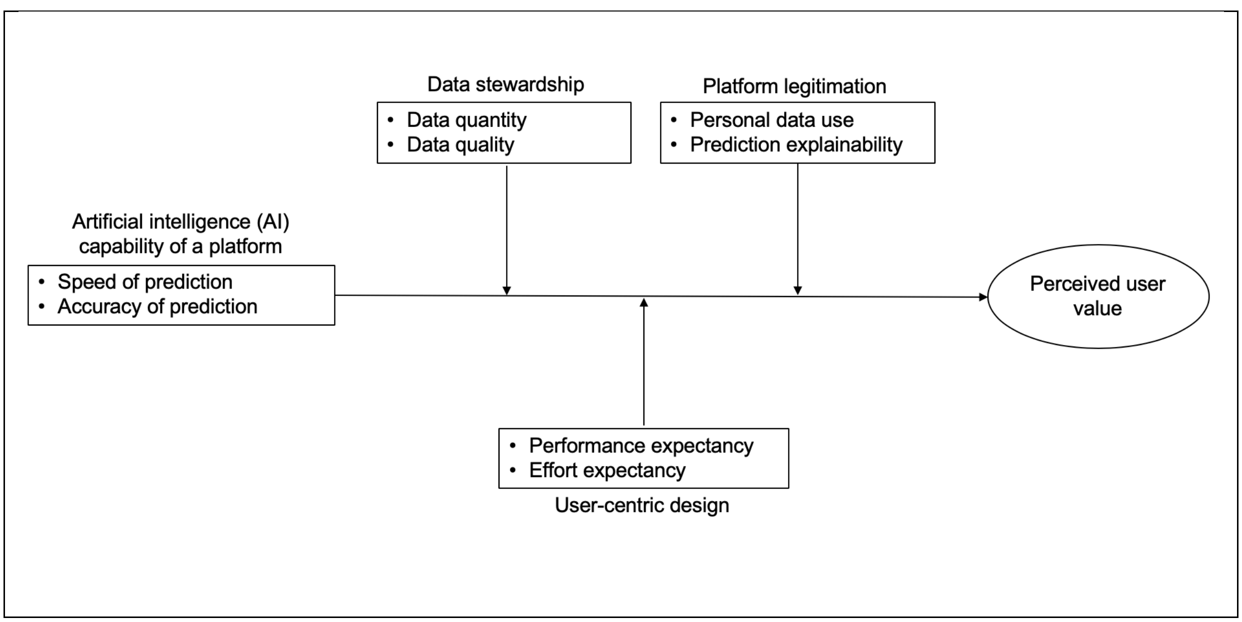

FIGURE 1: Model of Data Network Effects

Author bios:

Robert Wayne Gregory is Associate Professor at McIntire School of Commerce, University of Virginia. He received his Ph.D. in Management Information Systems from Goethe University Frankfurt. His research interests include digital innovation & strategy, platforms, and technology-driven change.

Ola Henfridsson is Professor of Business Technology at Miami Herbert Business School, University of Miami. Ola is also appointed as a WBS Distinguished Research Environment Professor at Warwick Business School in the UK and a KIN Fellow at VU Amsterdam. His research interests relate to digital innovation, platforms, and technology management.

Evgeny Kaganer is Professor and Dean for Academic Affairs at the Moscow School of Management SKOLKOVO. He is also a visiting professor at IESE Business School, University of Navarra. He received his Ph.D. from Louisiana State University. His research interests focus on digital strategy and organizational transformation.

Harris Kyriakou is Assistant Professor of Information Systems at IESE Business School, University of Navarra, and a member of European Commission's expert group on digitization. He received his Ph.D. from Stevens Institute of Technology. His research focuses on the intersection between collective and artificial intelligence.