From Open to Balkanized Data Flows in Digital Platforms: Toward a Framework of the Paradox of Data Openness in the Age of Algorithms

Likoebe M. Maruping
J. Mack Robinson College of Business
Georgia State University
LMaruping@gsu.edu

Yukun Yang
College of Business
Southern University of Science and Technology
yukunyang.cis@gmail.com

Forthcoming in *Strategic Management Review*

Abstract

The emergence of artificial intelligence and the need to train robust models has elevated the value of data generated on digital platforms. This has precipitated growing efforts to restrict or entirely cut off access to data generated on platforms. While the literature over the past decade has emphasized the steady progression toward increasingly open data flows for value creation in open collaboration, much less attention has been devoted to theorizing the ramifications of increasingly restrictive data flows, leaving little guidance on how strategic management can interpret and navigate platform value within these emerging trends. In this research, we develop a framework that offers guidance on how decisions to open or restrict data flows can affect value creation and value capture by key platform participants and the ecosystem as a whole in open collaboration. Considering the nature of the data as being asymptotic versus non-asymptotic, we theorize the paradox of data openness regarding value creation and value capture under different data openness or restrictiveness regimes. The framework serves as a basis for a variety of future research questions regarding the future of platforms in an AI and data landscape.

Keywords: platforms, artificial intelligence, data restriction, closed data flows, value creation, value capture, open collaboration, paradox

Acknowledgement: We gratefully acknowledge the guidance and feedback of Arvind Malhotra, Fabrice Lumineau and participants in the special issue workshop. We also express our gratitude to the anonymous reviewers and participants in the University of Queensland Business School brownbag seminar for constructive feedback that helped us strengthen the ideas expressed in this article.

Introduction

Digital platform ecosystems—comprised of digital platforms and participants—have been the prime beneficiaries of the steady progression toward openness in collaboration for value creation (Adner & Kapoor, 2010; Kretschmer et al., 2022). Open collaboration in digital platform ecosystems is conceptualized as a system of innovation or production where goal-oriented yet loosely coordinated participants interact to create a product (or service) of economic value (Levine & Prietula, 2014). In this system, complementors and users play critical roles by actively participating on digital platforms. Users engage in platform activities that generate valuable data, while complementors leverage this data to create valuable products and services that benefit the user base (Engert et al., 2025; Liu et al., 2024). Meanwhile, the platform owner acquires value created by both users and complementors through their sustained engagement. In this open collaboration, users, complementors, and platform owners emerge as essential participants, collectively shaping the value proposition of the digital platform ecosystem.

The push toward openness in collaboration to facilitate value creation has been constant, evolving from its inception in dyadic inter-organizational systems (e.g., Choudhury, 1997) to interfirm collaboration networks (e.g., Burford et al., 2022) to realizing its full expression in digital platforms (e.g., Parker et al., 2017; Rietveld et al., 2019). Although specific business models vary, a common thread is that digital platforms embrace openness in the participation of complementors and openness in data flows. By openness in data flows, we mean the free movement of data generated during activities on digital platforms due to being openly available for access, collection and storage to the parties involved (Ghazawneh & Henfridsson, 2013). The impetus for this drive toward ever-greater openness is that it enables digital platform ecosystems to continuously scale value creation and capture (Jacobides et al., 2018; McIntyre & Srinivasan, 2017; Parker & van Alstyne, 2018). Digital technology infrastructure has been a core enabler of this evolution of interfirm collaboration from dyads to networks to platforms (Adner & Kapoor, 2010; Constantinides et al., 2018). These technologies enable efficient data sharing between collaborating parties (e.g., Ceccagnoli et al., 2012). Today, application programming interfaces (APIs) make it possible for complementors to build their own value-creating services that operate on top of data generated on digital platforms (Parker et al., 2017; Schreieck et al., 2016; Xue et al., 2019).

If the 2010s to 2020s have been about openness seeing its full expression in digital platform ecosystems (e.g., Parker & Van Alstyne, 2018; Zhu et al., 2006), current indications are that the emergence of intelligent agents—software algorithms that learn from data and are capable of acting of their own accord—spells a potential moderation or even pull back of this trend (Krämer & Shekhar, 2025; van der Vlist et al., 2024). Two primary considerations underlie this development. First, robust digital infrastructures have made it possible for digital platforms to amass data assets of unprecedented scale in structured and unstructured forms (Nambisan et al., 2017; Tilson et al., 2010). These data assets, constantly being updated through ongoing participant activities, are a major source of competitive advantage for digital platforms (Baesens et al., 2016). Second, intelligent digital technologies are becoming a major source of new value creation (e.g., Gregory et al., 2021) and they rely principally on large-scale datasets to learn and to improve their performance (Raisch & Krakowski, 2021). This suggests that there is a distinctive advantage to the accumulation of large-scale data assets and generating user value through data network effects (Gregory et al., 2021).

This development, regarding the distinctive value of data, presents new challenges for open collaboration in digital platforms that existing literature has not directly addressed. For digital platforms, maintaining openness in data flows raises the possibility that various actors will have access to the same data assets on which to train the models behind their intelligent agents. Most generative foundation models (GFMs), such as BERT and OpenAI's GPT family of language models, are trained on common, publicly available data sets (Huang & Siddarth, 2023). In contrast, proprietary data offers the potential for competitive advantages in model training. For instance, BloombergGPT, which is trained on Bloomberg's extensive financial data, has been found to outperform ChatGPT across a range of finance-related natural language processing tasks (Wu et al., 2023). Proprietary models improve performance in particular use cases based on the proprietary data sets on which they have been fine-tuned (Lei et al., 2023). Thus, maintaining openness raises the risk of forgoing a substantial value creation opportunity for the platform owner (Gregory et al., 2021). Conversely, closing off and protecting data flows risks eroding the value that complementors can bring to the digital platform ecosystem, and may lead some to leave for competing digital platforms that are more open or could lead them to shut down. As an example of this possibility, Apollo, a popular complementor app for browsing Reddit on mobile devices, shut down operations due to Reddit's decision to start charging for API access (Peters, 2023). The choice of whether to open or restrict data flows presents a conundrum for digital platform owners and there are limits to the guidance offered by extant literature.

There have been notable shifts among some digital platforms in their stance toward openness in data flows. X (formerly Twitter) has begun charging for large-scale access to its data via the X API (CNBC, 2023) in order to generate revenue from its data asset. Similarly, Reddit, recognizing the value of its large corpus of data, started charging companies for access to its API, citing concerns that they might otherwise use that data for free to train their own chatbots (Issac, 2023). Recently, Reddit inked a $60 million per year licensing deal with Google, allowing Google to train its AI on user-generated content on the platform (Tong et al., 2024). Question and answer platform Stack Overflow has followed suit, charging companies for access to its large corpus of data for training intelligent agents (Dave, 2023).

These developments suggest that digital platform ecosystems may be on the verge of an evolution from open to Balkanized data flows, characterized by fragmented and siloed data-sharing practices in digital platforms, where data access is restricted to specific participants, often based on proprietary interests or competitive strategies. This shift has significant implications for open collaboration by participants in digital platform ecosystems, including platform owners, complementors, and users, as well as business models for value creation and capture. Strategic management research has developed robust frameworks and conducted extensive empirical investigations of collaboration through value creation and capture in open environments. However, it is important to consider what these recent trends portend for strategic management scholarship on open collaboration. There are emerging need and opportunity to inform our understanding of the evolving landscape for such collaboration. This motivates our core research question: *What are the value implications of opening versus restricting data flows in digital platforms?*

The purpose of this research is to address this research question by developing a framework that can serve as a basis for theorizing the potential paradox of data openness in terms of value creation and value capture among participants within digital platform ecosystems under conditions of Balkanized data landscapes. The anticipated value proposition of the framework and the article as a whole is threefold. First, it will challenge core underlying assumptions regarding the openness of data flows that have been the basis for the discipline's theorizing over the past decade, encouraging strategic management scholars to adapt their theories and to theorize accordingly (Bercovitz & Chesbrough, 2020). Second, it will serve as an initial platform for strategic management scholarship to develop new theories and empirical investigations of collaboration in a data-segmented competitive landscape. Third, it will provide an ecosystem-level view of how data openness decisions and the development of machine learning models potentially impact competitive landscapes in platform environments.

Our proposed framework is grounded in the paradox of openness to surface the tensions that emerge at the intersection between the free flow of data on digital platforms and open collaboration by participants in the platform ecosystem. The framework will draw attention to the global (ecosystem-wide) and local (participant-specific) tensions that may be evoked by enacting restrictions on data flows. In the sections that follow, we first overview the paradox of openness, which serves to ground our problem formulation and proposed framework. Second, we review and synthesize the literature on the openness of digital platform ecosystems, specifically focusing on the evolution of data-sharing agreements on digital platforms. Third, we begin to delve into the conceptual scaffolding for our framework by identifying the role and nature of data as a digital asset and discussing its implications for the value of data in openness strategies. Fourth, through the lens of the paradox of openness, we theorize the implications of various data openness-restriction regimes for global and local value creation and capture by participants in digital platform ecosystems. We conclude by proposing digital platform strategies for facilitating, monitoring, and controlling data openness so that the digital platform can maintain a healthy and open collaborative relationship with complementors and users while at the same time capturing value generated from its data assets.

Taken together, the theoretical framework that we propose to develop will contribute to strategic management scholarship by providing a lens through which to make sense of the evolving data landscape and its implications for navigating the tensions of data value creation and value capture in digital platform ecosystems.

Background and Literature

Paradox of Openness

Smith and Lewis define paradox as “contradictory yet interrelated elements that exist simultaneously and persist over time” (2011, p. 382). Paradox theory has been studied extensively in organizational and strategic contexts, as a way of highlighting the complex tensions that organizations face when managing competing demands (Smith & Lewis, 2011). At its core, paradox entails an understanding that seemingly individually logical elements can simultaneously be in contradiction with each other when juxtaposed and that any responses related to these elements must embrace the underlying tensions between them (Smith & Lewis, 2011). Among the various paradoxes that have been identified in the literature, we focus on the paradox of openness, given its connection to our emphasis on openness in data flows. Laursen and Salter eloquently articulate this paradox in the context of innovation by noting that firms must be "“open” by engaging with a broad set of external actors in their innovation activities but also have to protect their own firm’s knowledge from being copied by competitors” (2014, p. 868). A digital platform is considered open when the platform owner eases the restrictions on the use of the resources on the platform, the development of complements to the platform, and the commercialization of platform technologies (Boudreau, 2010).

In general, a digital platform can be opened in three ways. *Opening access to the platform* happens when the platform owner allows external users and complementors to participate and use the resources on the platform (Alexy et al., 2018; Eisenmann et al., 2009; O'Mahony & Karp, 2022) while the platform owner retains decision rights (Kretschmer et al., 2022). For example, the platform can provide interfaces (i.e., API) to allow outside developers to create complementary applications on top of the platform (Karhu et al., 2018). *Opening resources of the* *platform* emphasizes opening the platform's valuable resources by relinquishing the intellectual property rights (IPR) of the resource. For example, Google made and maintains the Android platform's core open source under the Android Open Source Project (AOSP) (Karhu et al., 2018). *Devolving control of the platform* refers to the right to determine the rules that guide the platform's usage and technical trajectory (Boudreau, 2010; O'Mahony & Karp, 2022). The platform owner offers development resources while retaining decision rights on leadership, while the participants contribute their creativity and resources, to which they retain at least partial control rights (Gawer & Cusumano, 2008; Kretschmer et al., 2022; O'Mahony & Karp, 2022). For example, Sun Microsystems made Java accessible to users under an open source license while retaining control over Java protocols (Garud et al., 2002; O'Mahony & Karp, 2022).

Digital platform owners are confronted with the need to keep the platform open so that complementors can use the open resources to create value for the ecosystem while at the same time facing pressure to keep it closed in order to keep competitors from appropriating value. On one hand, opening the platform can decrease the costs of producing a bundle of resources for the merit of the scalability of digital platforms (Alexy et al., 2018; Chesbrough, 2003; von Hippel & von Krogh, 2003), spur innovation on the platform by leveraging the wisdom of crowds (Baldwin & von Hippel, 2011; Kretschmer et al., 2022), better predict demand for the platform's services (Seidel et al., 2017) and grow a more extensive base of users (Wormald et al., 2023). On the other hand, data privacy and security have emerged as critical risks when opening the platform, increasing transaction costs for platforms (Kretschmer et al., 2022; Say & Vasudeva, 2020). Another concern with opening the platform is the exposure of intellectual property to rivals (Farrell & Simcoe, 2012; O'Mahony & Karp, 2022), which may lead to the forking of core technologies or intensifying competition (Parker & van Alstyne, 2018). Growth in complementors may make it more challenging to ensure the coordination and completion of value creation activities, considering the heterogeneity of complementors on the platform (Kretschmer et al., 2022; Zhang et al., 2022). Excessive openness may also incur "Atari shock"—an influx of low-quality complements that harms the benefits derived by ordinary participants (Kretschmer et al., 2022; Panico & Cennamo, 2022). Research on open innovation has noted that in order to facilitate and scale innovation, platforms must remain open, yet at the same time, keeping closed enables them to protect the intellectual property behind those innovations (Parker & van Alstyne, 2018; Felin & Zenger, 2020). Lin and Maruping (2022) have theorized how digital startups manage this paradox when innovating new products on open source platforms.

Strategic management research has explored the implications of openness in innovation (e.g., Laursen & Salter, 2020; Felin & Zenger, 2020; Sha & Nagle, 2019), conceptualizing open innovation as a strategic decision process (Chesbrough & Tucci, 2020). Specifically, strategic management research views open innovation as being concerned with how to allocate resources, how to organize, and how to win in the marketplace, and therefore simultaneously touches on topics such as strategic decisions, business models and competitive advantage (Bercovitz & Chesbrough, 2020; Chesbrough & Tucci, 2020). However, these and other open innovation studies primarily focus on inflows and outflows of knowledge and technologies as the main resources whose access should be open or restricted. In contrast, our study emphasizes data as a new resource whose openness or restriction should be a strategic consideration. In the next section, we describe data as a new form of digital asset that has been overlooked in prior literature and compare it to knowledge and technologies.

Data as a Digital Asset

Data have been identified as a critical resource on digital platforms (Alaimo & Kallinikos, 2022; Karhu et al., 2018). Alaimo and Kallinikos define data objects as "structured entities with a lifespan composed of aggregated data, which are organized according to a logic or schema" (2022, p. 29). When accumulated, data are editable, programmable and sharable in ways that can be tailored to be valuable to various participants. In a digital platform ecosystem, activities of the platform owner, users, complementors, products, and services can be rendered as data objects that can be accumulated for use in other value creation activities, principally through algorithms (Baskerville et al., 2020).

Every participant in a digital platform ecosystem can derive and appropriate data value generated in the platform. For platform owners, participant data provides deep insights into the networks of users and complementors. Product and service data helps identify emerging trends and gaps in the market (Chen et al.,2012). Feeding activity data into machine learning models improves platform functionality in optimizing user experiences, enhancing product recommendations, and streamlining operations. Gregory et al. (2021) have identified data network effects as another value proposition of data. Data network effects manifest when the more the platform learns from the data collected on its participants, the more valuable it becomes to each participant (Gregory et al., 2021). This development recognizes the increasingly central role of data (generated from user activity on digital platforms) in generating value on digital platforms. For example, Tesla engineered its cars to continuously collect accurate driving data. The data are used to troubleshoot problems, improve car performance, train and calibrate its artificial intelligence (AI) components, and send updates back to all cars (Niculescu et al., 2018). It turns out that the more Tesla drivers there are on the road, the better the car performance becomes, which benefits every driver (Niculescu et al., 2018).

Data is a unique type of platform resource compared to other resources such as knowledge and technologies that have been well studied by prior open innovation research (Teece, 2020). First, data are unstructured digital records generated from multiple sources and need to be processed and analyzed (through AI model training, as an example) in order to create meaningful insights and generate value. In contrast, technologies are structured and well codified tools and algorithms that are leveraged by actors in open innovation. Second, data are usually raw, unprocessed digital records, while knowledge is interpreted as understanding, and is often in the form of tacit, experiential and context-dependent expertise possessed by actors on platforms. It is not as fluid and malleable as data. These differences make the value assessment of data challenging and the control of data difficult. Although opening knowledge and technologies has been studied extensively by existing open innovation research, data openness has not received as much attention. Considering the potential value generated from analyzing data, especially in light of the adoption of AI-related technologies, the strategic value of data has been increasingly recognized by platforms and its actors.

To facilitate data value generation from various participants, platform owners can open access to their data to encourage collaborations and amplify data network effects. In a subsequent dialogue, Clough and Wu (2022) refine the theory of data network effects by introducing two complementary insights. First, they argue that data serves as an internal asset of the platform, creating a loosely coupled relationship between the quantity and quality of platform-owned data with the installed base of users and complementors. We build on this by emphasizing that while the platform owner maintains ownership of data, complementors are granted access to these data assets. Additionally, our framework also echoes Clough and Wu's (2022) second point that value capture, alongside value creation, is an essential consideration in platform dynamics. This idea is reflected in our framework as we explore the tensions between value creation and value capture, highlighting them as central to the paradox of data openness. Furthermore, the growing role of AI in value creation from data underscores the importance of managing diverse participant interests to navigate these tensions (Gregory et al., 2021). In the next section, we discuss the role of AI in shaping data openness strategies, hence bringing data as a strategic resource into strategic management research.

AI and Data Openness

As alluded to in the introduction, there appears to be a recognition of the value proposition of data, and platform owners are increasingly looking to develop their own proprietary AI models based on the data generated on their platforms to create value for platform participants. This shift in focus indicates that an appreciation for how to approach the openness of data assets has become a strategic imperative, as data now constitutes the raw material for building dynamic capabilities and achieving competitive advantage in the era of AI. For instance, Salesforce—a major customer relationship management platform—has developed and trained Einstein GPT[1], its own AI model (trained on data generated on the platform) that works in conjunction with OpenAI's ChatGPT to help complementors identify sales opportunities, build client relationships, and deliver personalized customer experiences. In delivering this AI-enabled value creation, Salesforce restricts complementors' access to customer data through its Einstein GPT Trust Layer, which is equipped with a number of data security guardrails such as data masking, in-flight encryption, and Zero Data Retention for external LLMs. This makes it possible for third parties to train their own models while retaining the data on the Salesforce platform. Another example is Bloomberg, which owns a platform for financial trading data and news and has developed and trained its own AI model—BloombergGPT. It has done so using a combination of FinPile, its proprietary dataset of archived data (such as news, financial filings, company press releases, social media, etc.), and data on the open internet (such as GitHub, ArXiv, Wikipedia, etc.). This model has been shown to outperform other LLMs in its responses

[1] See https://www.salesforce.com/products/einstein-ai-solutions/?d=cta-body-promo-8, accessed on Mar 29, 2024.

to finance-related natural language prompts (Wu et al., 2023). Bloomberg allows users to interact with the model through its terminals but restricts access to the training data as this is a highly valuable asset.

These examples illustrate the evolving landscape of how platforms are adapting to the emergence of platform-generated data and AI for value creation and appropriation as they consider what this all means for their own competitive position. Unfortunately, as this is a recent phenomenon and one that will likely evolve, there is not much by way of theoretical frameworks to help make sense of these developments and what they potentially portend for platforms and the participants of which they are comprised. As exemplified by the consequences of Reddit's decision to monetize access to its API, the ramifications can be wide-ranging, including complementors shutting down operations—as in the example of Apollo, BeaconReader and Sync—and complementors having to charge their users in ways that reduce their user base.

This research focuses on data openness as a strategic response to the challenges posed by the massive data accumulation and AI-related advancements within platform ecosystems. Specifically, we define *granting data openness* as the platform's provision of interfaces to enable external complementors to access and utilize the data assets while the platform retains full ownership and control over the data, including the authority to establish and enforce protocols governing the use of data. In the next section, we develop a framework that further discusses the paradox of data openness, manifesting in tensions between value creation and value capture by different participants.

A Framework of the Paradox of Data Openness

In light of our review of the literature on digital platform ecosystems, the promise and risk of openness in such ecosystems, and value of data as a digital asset in such ecosystems, we now develop a conceptual framework. The framework recognizes that digital platform ecosystems comprise multiple, loosely connected participants who have different motives but engage on the digital platform to create and capture value. The framework positions the platform owner as the entity who makes decisions that determine what participants can do within the ecosystem. Our theorizing about the decisions of the platform owner is grounded in the paradox of openness. The premise underlying the framework is that platform owners confront a paradox of data openness whereby granting data openness is necessary to facilitate value creation and capture by various participants in the ecosystem, while at the same time restricting data openness is necessary to protect the platform owner's data asset from appropriation that can erode its competitive position.

The crux of our framework is that platform owners need to adopt the appropriate data openness regime given the nature of the data involved, and these choices will evoke or alleviate tensions for particular participants within the ecosystem. Before we elaborate on the framework, we briefly introduce the nature of data—in terms of its temporality—that is essential to understanding the value creation and capture tensions and how they might be alleviated. We emphasize the implications of the accumulation and temporality of data because the training of proprietary machine learning models that operate at scale on digital platforms relies on large volumes of data (Agrawal et al., 2018). We first attend to the conceptualization of the temporality of data. Having established this important conceptual foundation, we next surface tensions with regard to key participants in a digital platform ecosystem—principally platform owners, users, and complementors. These constitute the most commonly considered participants in digital platform research (e.g., Boudreau & Jeppesen, 2015; Kretschmer et al., 2022). We develop our framework of the paradox of data openness in light of the temporality of data and decisions to restrict or open access.

The Temporality of Data

As already noted, data network effects represent a highly consequential mechanism for digital platforms to generate value from data (Gregory et al., 2021). However, we believe that the nature of value inherent in the data produced from participant activities on digital platforms has important implications for value creation and value capture. Of particular relevance for our purposes is the *temporality* of data—the time dependent nature of its value. Data differ in their temporality. Some data maintain the same value over time because their meaning is not tied to temporally sensitive events. Other data are more time dependent and experience a decay in value as time passes (similar to the notion of a half-life). We refer to the former as *asymptotic* and the latter as *non-asymptotic*. Note that the temporal nature of data is distinct from the accumulation of data. Accumulation is a matter of the quantity of data. The value implications of the accumulation of data can be quite different depending on its asymptotic versus non-asymptotic nature, as we will describe next. Table 1 summarizes the definitions.

Table 1. Definitions of the Nature of Data and its Value

Concept	Definition
Data temporality	The degree to which the value of data is time dependent.
Asymptotic data	Data whose accumulation is subject to diminishing returns in value over time.
Non-asymptotic data	Data whose accumulation consistently adds value over time

Asymptotic data are data for which incremental accumulation adds value at a progressively decreasing rate (Currier, 2020). After large volumes of data have been accumulated, every incremental amount of data accumulated asymptotically approaches a ceiling in terms of additional value. One example of this is data used to identify medical conditions through classification by machine learning models. The value of these data is not time dependent. The accumulation of the first 1 million data points likely significantly improves the prediction accuracy of the machine learning model. However, the 1-million-and-1st additional data point likely contributes less to improving accuracy and the 1-million-and-2nd additional data point even less. In essence, the data are subject to diminishing returns with regard to cumulative value.

Non-asymptotic data are data for which accumulation adds value at a near-constant rate. Every time new data accumulate, they add greater value in prediction accuracy than older data points (Currier, 2020). This occurs in situations where the value of data is context or time dependent. For instance, the algorithm on the Uber app better produces accurate pickup wait times using data on current traffic conditions in a user's location relative to using accumulated data from three years ago in the same location. With such data, the ongoing accumulation of data is less subject to diminishing returns because its value is dependent on its recency.

Consequently, the temporal nature of data has implications for data network effects derived therefrom. It suggests that there may be differences in the time value of data accumulation.

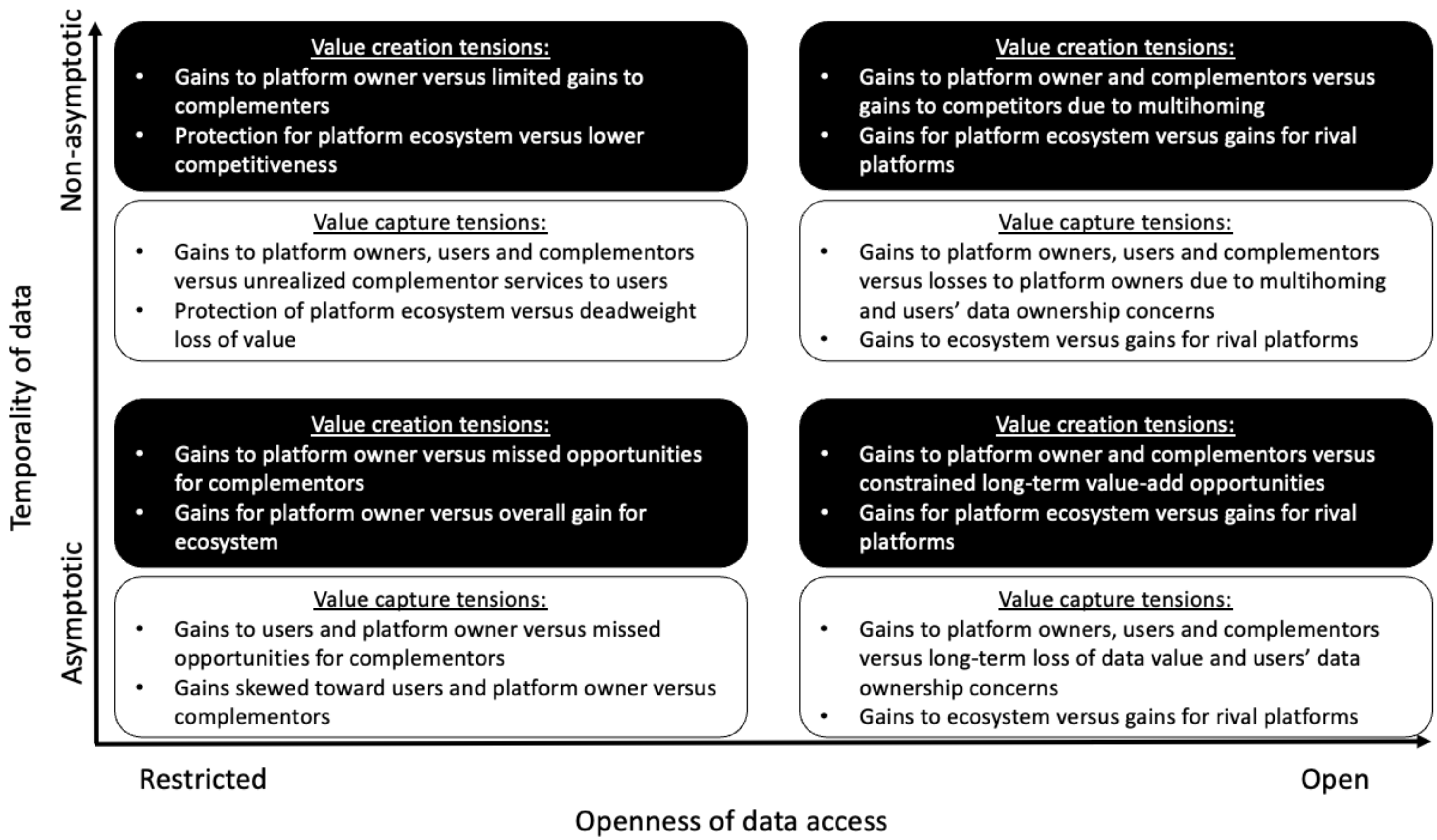

Figure 1. Paradox of Data Openness Framework in Light of Data Temporality and Access

Restricted Access to Asymptotic Data

Accumulation of asymptotic data emphasizes the ongoing acquisition and storage of vast quantities of data from participant activity on a digital platform. Here we emphasize the main value proposition of such data as encompassing its volume. Digital platforms that facilitate standardized, repeatable interactions (e.g., buying and selling of goods, consumption of a service) are known to accumulate data on stable attributes such as order quantities and user choices that occur with some regularity over time (Lei et al., 2023). Gregory et al. (2021) point out that the accumulation of such data in large volumes is valuable for training machine learning models. Specifically, they point out that having a larger number of past cases facilitates greater accuracy in prediction by machine learning models. Indeed, a common feature of developing robust machine learning models involves ensuring that they are trained on large volumes of data (Agrawal et al., 2018). It is a large part of the reason why big data was an important precursor to the re-emergence of AI (Agrawal et al., 2018). Further, machine learning models trained on large volumes of past data can be fine-tuned over time through reinforcement learning using ongoing participant feedback and activity on the platform. Consequently, volume of accumulated asymptotic data is a useful attribute for evaluating potential tensions that can emerge when access to such data is restricted by the platform owner. As discussed earlier, platforms are able to control access to data generated on the platform through technology interfaces—primarily in the form of APIs (Ghazawneh & Henfridsson, 2013).

Tensions in Value Creation. Restricting access to accumulated asymptotic data evokes notable tensions across participants and for the digital platform ecosystem as a whole. First, it skews value creation opportunities toward the platform owner at the expense of complementors.

By restricting access to asymptotic data, the platform owner positions itself to create value through developing proprietary machine learning models. These large volumes of transaction data enable the platform owner to develop machine learning models with high levels of accuracy (Agrawal et al., 2018). Moreover, because they are trained on proprietary data, they have the potential to enhance the platform owner's value proposition to users. For instance, Amazon uses trained models to aid sellers in its marketplace in setting their product prices to maximize sales (Chen et al., 2016). Such data-enabled value creation facilitates data network effects—enhanced platform-specific services make the digital platform attractive to users, and more users generate additional data that can be used to fine-tune the algorithms and so on (Gregory et al., 2021). At the same time, however, restricting data access impacts the ability of complementors in the platform ecosystem to develop complementary products and services that contribute to value creation in the ecosystem (Hukal et al., 2020). To be sure, complementors should still be able to produce complements to the digital platform's main offering under such a regime. However, their restricted access to the digital platform's large volume of asymptotic data will likely constrain their ability to train robust machine learning models that are fine-tuned to the needs of the digital platform's users. Additionally, because of the asymptotic nature of the data and model training, over time platform owners will realize lower marginal gains in value creation as more such data contribute less to model performance. Consequently, at some point the value of restricting such data for developing its own robust proprietary machine learning models will be outpaced by the foregone potential value creation by complementors if access were granted (Parker & van Alstyne, 2018).

Second, restricting access to asymptotic data enhances value creation locally while potentially being suboptimal for the ecosystem as a whole. As noted above, restricted access enhances the platform owner's ability to deliver better services to users of the digital platform, enhancing direct network effects. However, the underlying tension of this arrangement is that it constrains complementors' ability to create value from those data. Complementors are generally in a better position to organically identify and pursue value creation opportunities that are complementary to the digital platform's offerings (Hukal et al., 2020; Parker et al., 2016). Those complementary offerings are responsible for generating indirect network effects that enable the ecosystem as a whole to create greater value.

In sum, restricted access to asymptotic data evokes tensions in value creation by opening up opportunities for the platform owner while constraining opportunities for complementors and expanding value creation potential for the platform owner while constraining it for the ecosystem as a whole.

Tensions in Value Capture. Noteworthy tensions emerge with respect to value capture by participants and welfare distribution across the ecosystem when access to asymptotic data is restricted. Among participants, users and platform owners are likely beneficiaries with regard to value capture. Users are most poised to capture the value created by machine learning enabled services on the digital platform (Huang et al., 2022). Parker et al. (2016) argue that algorithms play an important role in ensuring that the right interactions that should happen, do happen. Deployment of more accurate machine learning models can help ensure that users are aware of, or directed to, on-platform interactions that will best meet their needs. Such precision can potentially make their time on the digital platform a more valuable experience. The platform owner is also poised to capture the value created by deploying models trained on proprietary data generated on the digital platform. Specifically, they are in a position to capture value from direct network effects. When users perceive the value from data network effects (Gregory et al., 2021), they are likely to engage in more interactions, which makes the platform more valuable to each user and increases the volume of activity from which the platform owner can capture value (e.g., through charging for access or charging a transaction fee) (McIntyre & Srinivasan, 2017).

Conversely, complementors are least likely to capture the value created by the deployment of machine learning models on the digital platform. Absent access to asymptotic data, they may not be in a position to tailor their offerings to meet user needs with great precision. For instance, when Apple Inc. implemented its app tracking transparency policy—which limited the ability to track Apple device user data—Meta argued that it would result in upwards of US$10 billion in lost sales in 2022 alone because advertisers on its Facebook platform would not be able to target their ads to users as accurately (Leswing, 2022). Consequently, the share of value that complementors are able to capture from users engaging with their offerings on-platform may be limited. Interestingly, users may also inadvertently fail to capture value from indirect network effects. Limitations in the value delivered by complementors means that users do not get to enjoy the expanded set of offerings that they otherwise would. Complementor offerings may not meet user needs with a desired degree of precision, hence eroding the amount of value they derive.

From the perspective of the digital platform ecosystem, restriction of access to asymptotic data is likely to skew the distribution of welfare toward users and the platform and away from complementors. At first glance, this appears to be good for the ecosystem as a greater share of the value is being captured by the users who generate the data through their activity streams on the platform. However, a potential risk to the digital platform ecosystem is that if complementors are not able to capture adequate value, they may leave or redirect their resources toward a competing digital platform (Cennamo & Santaló, 2013; Loh & Kretschmer, 2023), which in turn can erode the platform's value to users.

In sum, users and platform owners stand to capture value from restricted access to asymptotic data, while complementors are less likely to capture that value. Further, at the ecosystem level, the distribution of gains in value from data network effects accrues to users and platform owners in ways that potentially threaten the value proposition for complementors. This entails potential risk to the viability of the ecosystem.

Open Access to Asymptotic Data

The central proposition of asymptotic data is its value generation through accumulating a large volume of activity data on the digital platform. As discussed earlier, the value of such data does not grow exponentially with its volume. Instead, its marginal value asymptotes to a stable level (Currier, 2020). Despite the nature of diminishing marginal value, platform owners may consider opening these data to complementors in order to expand value creation opportunities.

Tensions in Value Creation. Opening access to asymptotic data benefits both platform owners and complementors. Platform owners can utilize asymptotic data as a reliable start to train their proprietary algorithms and provide value to their users. However, the value that platform owners acquire from the asymptotic data is limited due to the progressively decreasing marginal value. On the one hand, machine learning models can be fed with a large accumulation of data, leading to a sharp increase in prediction accuracy. On the other hand, as such data cannot offer constant value over time (Currier, 2020), platform owners may find themselves unable to maintain the value-add for their users, thus weakening the platform's competitiveness.

Providing access to complementors creates opportunities to develop their products on the basis of the platform's value. With the platform's support, in the form of function libraries and

APIs, complementors can combine their products with the platform's core value proposition (Niculescu et al., 2018). In this way, complementors and platform owners collaborate to expand the value they can offer to platform users (Karhu et al., 2018). These collaborations also slow down data aging and delay the point at which value hits diminishing returns.

However, opening access to asymptotic data has natural limits to value creation. As these data are relatively static, opening asymptotic data access may filter out complementors needing real-time data. For complementors who build their products on asymptotic data, the constrained growth of marginal value becomes a long-term concern. Additionally, opening access to such data means that competing platforms can access the data to train their machine learning models and offer their own competing machine learning-enabled services.

In sum, opening access to asymptotic data triggers tensions in value creation by opening up opportunities for the platform owner and complementors to seek value-adding collaborations while also constraining long-term growth opportunities for both the platform owner and complementors.

Tensions in Value Capture. The value associated with asymptotic data can be captured by platform owners, users, and complementors. Platform owners can capture the value complementors create in their continuous contribution and enrichment to the platform (Gregory et al., 2021). Users consume the value by enjoying the core products and services offered by the platform owner and the complementary products provided by complementors. Through collaborating with the digital platform, complementors can save costs on growing the user base and capture value when platform users consume their products (Niculescu et al., 2018). Taken together, all participants can enjoy part of the value created on the platform, which promotes a healthy and sustainable ecosystem.

Despite the benefits, platform owners and complementors may find it difficult to scale up their business even when asymptotic data are made accessible. Due to the nature of asymptotic data, most of the valuable insights are gleaned and offered to users at an early stage. The accumulation of data may confirm the insights while offering little additional value (Currier, 2020). Moreover, data ownership continues to be a critical concern for users. They may disagree with platform owners using or sharing their personal information or activity logs.

In sum, platform owners, users, and complementors are likely to capture value from opening access to asymptotic data, contributing to the value proposition of the whole platform ecosystem. However, there are constraints for both the platform owner and complementors in growing their business if their products and services largely depend on asymptotic data.

Restricted Access to Non-Asymptotic Data

Accumulation of non-asymptotic data emphasizes the acquisition of timely data regarding participants' activities on a digital platform. The main value proposition here is the timeliness of the data. As part of their conceptualization of data quality, Gregory et al. (2021) highlight timeliness—the speed with which data are updated as events occur—as a key attribute. As discussed earlier, timeliness is a meaningful attribute here because the accumulation of current data adds greater value to machine learning model improvement than do older data. Predictions are made based on present conditions rather than on past conditions. Hence, machine learning models trained and fine-tuned on such data have a high potential to add meaningful value to the participant experience on digital platforms. Uber and Waze are good examples of digital platforms that make use of non-asymptotic data to generate data network effects. Waze incorporates user data on current traffic conditions to fine-tune its predictive model and provide timely and accurate driving directions; this enables the Waze platform to create more value for its users and attracts other users to the platform (who themselves generate current data that can be used to fine-tune the machine learning model and so on).

Tensions in Value Creation. As with the case of restricting access to asymptotic data, restricting access to non-asymptotic data evokes tensions across participants as opportunities to create value through machine learning models trained on those data accrue exclusively to the platform owner. In contrast to the case of restricted access to asymptotic data, the value creation gains from non-asymptotic data do not erode with the ongoing accumulation of data from user activity. As such, the platform owner is in a position to perpetually create value through proprietary machine learning models trained on such data. Although complementors are still in a position to create value on the digital platform, they are not afforded the opportunity to create value through the use of non-asymptotic data for their own machine learning model development. A drawback of such restriction is that while it expands opportunities for value creation for users on the digital platform, it creates natural bottlenecks as that value creation occurs exclusively within the platform owner. This foregoes the opportunity of scaling machine learning model innovations based on such data externally, which is known to be more efficacious (Parker et al., 2017) and complementary to internal innovation efforts (Cassiman & Veugelers, 2006).

There are benefits to a platform ecosystem when access to non-asymptotic data is restricted. First, the platform owner is able to exert greater control over what machine learning-enabled services are developed and deployed on the digital platform. From a platform governance perspective, this can be beneficial in fostering a coherent image and experience for the platform (Chen et al., 2022). For instance, Google Maps restricts access to its data by charging its API on a pay-as-you-go basis. With the proprietary data, Google Maps is able to train its own AI models to add exclusive features (e.g., Immersive View, which offers a 3D view and tidbits of information of locations) to enhance the usability of its existing features (e.g., search for available EV charging stations for drivers), which improves user stickiness (Hawkins, 2023). Second, restricting access prevents complementors from multihoming with a value proposition based on data generated on the digital platform. Any value created on the basis of such data stays within the platform ecosystem. Third, restricting access to such data prevents rival platforms from accessing that data to develop their own competing machine learning models. From the perspective of the platform ecosystem, restricting access to non-asymptotic data constrains value creation opportunities for specific participants. While the ecosystem can realize gains in value creation under such a regime, for reasons discussed earlier, the scope of that value creation within the ecosystem will be constrained. One possible consequence of this is that it could make the platform ecosystem less competitive relative to rival digital platforms, hence attracting fewer users and complementors to the platform than might otherwise be the case.

Tensions in Value Capture. Restricting access to non-asymptotic data facilitates value capture for platform owners, users and complementors. Users directly capture value from very high-accuracy services based on up-to-date data that makes it easier to engage in value adding-interactions on the digital platform (Gregory et al., 2021; Parker et al., 2016). This enhanced value capture will both entice more activity from existing users and attract new users to the platform. Platform owners are able to capture the value from stronger direct network effects that are generated from a greater volume of high-quality interactions between users on the digital platform. This promotes a virtuous cycle in which increased user activity generates more timely non-asymptotic data that are used to further enhance machine learning model performance

(Currier, 2020). Complementors also capture some of the value that is created by platform owners' proprietary machine learning-enabled services. One way they capture this value is from the increased number of users and interactions on the digital platform. Another way they potentially benefit is when these services direct users to their complementary offerings on the digital platform.

There are several potential drawbacks to value capture in this arrangement. For users, while they clearly capture value from the platform owner's proprietary machine learning models, restricted access means that they forego similarly attuned machine learning-enabled services from complementors. This constitutes a consumer deadweight loss in economic terms because of the market's inefficiency in meeting users' needs when the capacity to meet those needs is available but unutilized (Hausman, 1981). That is, users are still able to capture value from indirect network effects generated by complementors, but that value is less than what would be realized if complementors are able to deploy their own proprietary models to cater to the unique needs of users. Complementors are able to capture value on the platform as already noted. However, they experience a producer deadweight loss because they are not able to capitalize on opportunities for user value due to data access restrictions (Hausman, 1981). Overall, under a regime of restricted access to non-asymptotic data, the platform ecosystem experiences deadweight loss—i.e., value that remains uncaptured.

In sum, restricting access to non-asymptotic data benefits platform owners, users and complementors but also constrains the magnitude of value captured by each of these participants on the digital platform.

Open Access to Non-Asymptotic Data

Tensions in Value Creation. Platform owners and complementors are likely to be beneficiaries of open access to non-asymptotic data. Platform owners benefit by being able to develop and train their own proprietary machine learning model to create value for their users. Non-asymptotic data enable platform owners to deploy machine learning models that deliver highly tailored services to users. Those models continue to be fine-tuned and improved as newer non-asymptotic data are accumulated from activity on the digital platform. By opening up access to such data, complementors also benefit in their value creation efforts. Such timely, high-quality data enables them to develop their own complementary machine learning-enabled services that are highly tailored to user needs, thus generating indirect network effects. Taken together, opening up access to non-asymptotic data increases opportunities for value creation on the digital platform.

Opening up access to non-asymptotic data can benefit the ecosystem as a whole. By distributing the opportunity to develop machine learning models that serve user needs, the prospects for scaling innovation for value creation increase. This is consistent with the observation by Parker et al. (2017) that innovation scales more effectively outside the boundaries of a firm than it does inside the firm. The variety and utility of machine learning-enabled services available in the digital platform ecosystem is likely to be more robust when multiple participants have an opportunity to innovate. This is consistent with Enkel et al.'s (2009) observation that inside-out approaches to innovation expand technological advancement by transferring ideas outside the firm. Indeed, Parker and van Alstyne (2018) argue that opening up access to intellectual property for innovation enhances the profitability of a platform ecosystem.

Specifically, the value of openness “rises with developer value-add and with resource reusability” (Parker & van Alstyne, 2018, p. 3025).

There are potential drawbacks to opening up access to non-asymptotic data. First, complementors with access to an ongoing stream of timely data may have an incentive to multihome. Once they have developed and trained their own proprietary machine learning models, they can deploy them to create value on competing platforms (Tian et al., 2022). Consequently, some value creation can end up accruing off platform. Second, opening up recent activity data for use by complementors can raise data ownership concerns by users. There is a growing chorus of concerns regarding the fairness of letting data harvester profit from users’ data (Zuboff, 2019). Third, opening up access to non-asymptotic data renders the platform ecosystem vulnerable to rivals who can access the data to develop their own competing services (Kretschmer et al., 2022).

In sum, there are considerable benefits for platform owners, complementors, and the ecosystem as a whole when opening up access to non-asymptotic data. It enables greater value creation through machine learning-enabled services that are highly tailored and sensitive to contextual user needs. At the same time, there are also drawbacks, including incentives for complementors to multihome, concerns from users about data stewardship, and opportunities for imitation by rival platforms who can also access the data.

Tensions in Value Capture. Platform owners, users and complementors are likely to benefit from capturing the value associated with non-asymptotic data. Platform owners capture the value from increased user activity that stems from context-specific services enabled by machine learning models. As these tailored services make for a more pleasurable interaction experience for users, the volume of interactions is likely to increase in the form of more interactions from existing users or more new users being drawn to the platform. Users, in turn, benefit directly from these machine learning-enabled services that provide highly contextualized guidance for their platform experience. Such personalization enhances the overall value users derive from their time on the platform. Complementors also benefit from non-asymptotic data when users make use of their machine learning-enabled services on the platform. Collectively, an open, vibrant, innovative ecosystem where all participants are capturing some value enhances the sustainability of that ecosystem. The non-asymptotic nature of data ensures that value continues to accrue over time, thus supporting the long-term sustainability of the digital platform ecosystem.

Despite its benefits, opening access to non-asymptotic data also entails drawbacks. First, platform owners may lose out on capturing some of the value if complementors decide to multihome their own machine learning-enabled services derived from these data. That may make other competing digital platforms attractive venues for users rather than joining the platform owner's digital platform. Second, users may raise data ownership concerns regarding who benefits from their data—particularly if the platform owner charges them for services derived from those data. For example, when GitHub released its paid subscription to Copilot—its machine learning-enabled service for generating software code based on natural language prompts—developers (users) on the platform filed a lawsuit arguing that the machine learning model was trained on code that millions of users had written (Roth, 2023). Third, the platform ecosystem could see its share of value capture erode if competing digital platforms use the openly accessible data to develop competing services.

In sum, opening access to non-asymptotic data enables platform owners, users and complementors to capture the aggregate value of such data. The digital platform ecosystem, as a whole, captures this value by drawing more participants to engage in its value proposition. However, this openness also introduces drawbacks. Platform owners may risk value erosion if complementors create services outside the platform or if users raise concerns regarding data ownership. At the ecosystem level, risks arise when rivals access and leverage these data to develop their own competing offerings, thereby threatening the ecosystem's competitive position.

Navigating Tensions in Value Creation and Value Capture

Our paradox of data openness framework highlights how platform owners' decisions regarding opening versus restricting the flow of data evoke tensions depending on the nature of the data at question. This raises a critical question: how can platform owners navigate these tensions? We attend to this question by first broadly articulating the nature of the underlying tensions that emerge from our framework. These tensions are best understood by decomposing the levels of analysis at which they are evoked. Specifically, tensions emerge both at the platform level and at the ecosystem level, for which we outline potential navigation strategies through a coopetition lens. Notably, we do not intend to provide an exhaustive list of strategies but rather to encourage future studies to explore more strategies for alleviating these tensions.

Navigating Platform-level Tensions

Within a platform—which comprises the platform owner, complementors and users—the fundamental tensions relate to the platform owner's need to safeguard the data asset from expropriation versus fully unlocking value from participant activity (Parker & Van Alstyne, 2018). In essence, platform owners are faced with navigating between enabling complementors to maximize value creation for users on the platform on the one hand—which is accomplished with greater openness in the flow of data—and ensuring that the value capture from such data flows accrues to all participants on the other. The extent to which these tensions are intensified or mitigated depends on platform owners' strategic orientation—whether leaning toward a more cooperative or a more preventative stance.

Cooperative strategies. Cooperative approaches to navigating the tensions entail prioritizing collaboration with complementors and distributing value more equitably among participants. Platform owners can adopt strategies that expand access to data while still safeguarding critical value. Cooperative strategies can involve monetary or policy levers for enactment. One such strategy is *priced access to data assets*. By offering tiered or usage-based pricing models, such as pay-per-use or subscription plans, platform owners can tailor access to complementors' needs, thus encouraging innovation while ensuring that the platform captures a fair share of the value created. It allows complementors to gain the data necessary to enhance user experience, while platform owners maintain control over how data is protected and monetized. This approach is particularly suitable for non-asymptotic data, as it encourages collaborative innovation on critical data assets while limiting the extent or frequency of access. This approach does not eliminate the possibility of expropriation, as complementors may still be able to use the data to create value off platform. However, it does ensure that the platform owner is capturing value in a monetary form.

A policy-focused cooperative strategy at the platform level is *opening access to preprocessed data*. In navigating the tensions between protecting proprietary data and fostering innovation, platform owners can preprocess raw data and open the "semi-finished" data product to complementors as a value-added service. Such an approach leaves data access open for value creation by complementors while safeguarding the integrity of the raw data. It preserves exclusivity over the most critical data while reducing deadweight loss caused by foregoing contributions from complementors. Such an approach does not completely eliminate deadweight loss, as complementors may be limited in the kind of value they can create with preprocessed data. However, it also limits expropriation as those same limitations to value creation apply outside the platform.

Preventative strategies. When platform owners adopt a more conservative stance—prioritizing the protection of their data and restricting complementors' ability to capture value at expense of other participants, they may employ *selective exposure* as a preventative hybrid measure. This strategy allows platform owners to determine what data are made accessible and to which complementors. In practice, most platforms manage a mix of data assets with both asymptotic and non-asymptotic properties. Selective exposure allows platform owners to align data access with the platform's value proposition. For example, open APIs can be configured to include only function calls to asymptotic data and paid access can be required for non-asymptotic data requests. In doing so, platform owners balance data openness with control, supporting complementor innovation while safeguarding their own competitive advantage. Such an approach provides platform owners with flexibility in how the tensions are navigated as it encompasses a configuration of monetary and policy measures.

Navigating Ecosystem-level Tensions

In contrast to the platform-level tensions, ecosystem-level tensions center on balancing the maximization of participation welfare within the ecosystem against the risk of value leakage to rival platforms. Two principal vectors for value leakage are of particular concern at this level: multihoming by complementors and direct access by rivals. Because complementors commonly participate across multiple platforms, value generated from data on one platform may ultimately benefit users on competing platforms, circumventing the ability of the focal platform's ability to capture that value. Similarly, rival platforms may directly access a platform owner’s data asset to create their own value to attract complementors and users. To address these tensions, platform owners can adopt two strategies, depending on their strategic orientation: a cooperative approach which emphasizes collaboration to expand the overall value created, or a preventative strategy which prioritizes restricting access to protect competitive advantage—even at the cost of limiting ecosystem growth.

Cooperative strategies. A cooperation-oriented strategy seeks to "grow the pie" by fostering complementarities rather than competition. As noted at the platform level, *opening access to preprocessed data* is also a suitable strategy at the ecosystem level, which enables broader innovation while safeguarding core assets. In addition, platform owners can also adopt the strategy of *differentiating data targets*. This approach directs complementors to focus on distinct aspects of data in building their competitive advantages, while platform owners retain control over data that are most critical to their core business. By steering complementors toward adjacent areas, platforms enhance collaborative innovation across the ecosystem while reducing direct competition arising from market overlap or resource similarity. Differentiating data targets thus fosters cooperation, expands the ecosystem's scope of value creation and secures value capture through clearly delineated domains of data demand.

Preventative strategies. As noted at the platform level, *selective exposure* can also serve as an effective preventative strategy at the ecosystem level. When the overarching goal is to prevent rival platforms from appropriating value from proprietary data assets, the focal platform owner may adopt a *strategic deferral* strategy. Rather than framing openness and restriction as a binary choice, platform owners can stagger access over time: initially restricting newer data while gradually opening older data. This approach can work particularly well for non-asymptotic data because recent data is key to machine learning model accuracy. Platform owners can retain exclusivity over newly generated data to strengthen their own models, while also giving complementors access to older data so that they can develop proprietary models to complement the digital platform's overall value proposition. Similarly, platform owners can make asymptotic data accessible to complementors once they have reached diminishing returns from fine-tuning models with accumulated data. Such deferral ensures that complementors are not locked out of opportunities to engage in value creation and value capture, thereby balancing competitive advantage with ecosystem health.

Another preventative strategy for navigating the tension is *selective encagement*. This policy measure entails encrypting high value data (such as data with asymptotic properties) so that it is only usable by complementors on-platform. Such data can be encrypted at-rest—i.e., when stored on platform servers. With such an approach, complementor products and services that rely on such data can only be realized within the platform—effectively engaging value creation and capture. Further, only authorized participants can decrypt such data using a platform-provided encryption key. A potential tradeoff with such an approach is that encryption slows down transaction performance on the platform due to additional computation necessary for encryption and decryption. This has potential to erode the consumption experience for users. Thus, efforts to lock in value capture on-platform, may come at the expense of value capture by users.

Discussion

This research set out to make sense of the evolving landscape for digital platform ecosystems where data have become a core aspect of delivering their value proposition. We observed that the emergence of intelligent algorithms has demonstrated the value of the data assets that digital platforms have amassed and that opportunities for value creation and capture would accrue to those that develop and deploy proprietary models that offer a distinctive value proposition. This development has prompted a rethinking of the virtues of openness that have guided platform thinking for the past decade. To make sense of this evolving landscape, this research developed a framework that theorizes the tensions that are likely to emerge when platform owners make the decision to restrict or open access to the data that are generated on their digital platforms. The framework identified the temporality of data and openness of data access as two axes along which to understand the tensions that are likely to emerge for specific participants and for the platform ecosystem as a whole.

Research Contributions

The framework developed herein contributes to the literature in several notable ways. First, our study contributes to research at the intersection of open innovation and strategy by reframing openness through the lens of data flows. Open innovation research has traditionally examined the inflows and outflows of knowledge and technology (Chesbrough 2003; Laursen & Salter, 2014; Shah & Nagle, 2019; Teece 2020). Data and its value have only recently begun to take center stage in open collaboration (Clough & Wu, 2022; Gregory et al., 2021). Consequently, it is little wonder that the literature lacks robust theoretical frameworks to make sense of this emergent turn of events. In this regard, this research contributes to strategy research by drawing attention to data assets as a strategic imperative and developing a theoretical framework to engage with the nexus of the nature of data and decisions to restrict or open access to it in digital platforms.

Second, this research advances our understanding of data network effects in digital platforms. Gregory et al. (2021) introduced the concept of data network effects and the role of AI capabilities built on data generated on digital platforms. However, their monolithic treatment of the nature of data risks masking the value creation and capture implications of data access decisions. We further refine the consideration of data network effects by drawing attention to the temporality of the data generated on digital platforms in the form of its asymptotic versus non-asymptotic nature. This nuanced view of the nature of data enables deeper theorizing about how digital platform owners can strategically manage data access to gain competitive advantages against others. Our differentiation regarding the temporal nature of data suggests that not all data network effects are created equal and should, therefore, not be theorized as such.

Third, our research contributes to the strategy literature by introducing the paradox of data openness as a novel lens for unraveling tensions in value creation and value capture in digital platforms. While Ryan-Charleton & Gnyawali (2025) draw attention to tensions between value creation at the firm level versus joint value creation, our framework moves beyond this focus by theorizing how openness of data flows and the temporal nature of data intensify or mitigate these tensions (Vasudeva et al., 2020). This expanded perspective highlights the multifaceted trade-offs that arise when platform owners restrict or open access to data: they can advantage some actors while disadvantaging others, and what appears beneficial locally may prove globally suboptimal for the overall platform ecosystem.

Fourth, our study contributes to coopetition research by shifting the focus from coopetition among complementors to coopetition between complementors and platform owners. Prior strategy research has largely examined coopetition among firms or among complementors in platform ecosystems, emphasizing how they simultaneously compete and cooperate in critical activities or markets (Chiambaretto et al., 2025; Vasudeva et al., 2020). Our work advances this line of work by demonstrating that coopetition also occurs between platform owners and complementors, especially as data emerges as a critical asset for which they strive (Minà &

Dagnino, 2025). On the one hand, both platform owners and complementors compete to train machine learning models using data generated on the platform, creating resource similarities and market overlap (Minà & Dagnino, 2025). On the other hand, they must cooperate to deliver valuable products and services for users. By foregrounding this paradoxical relationship, our research identifies a distinct form of coopetition that becomes especially salient in emerging phenomena, such as digital platforms, ecosystems, and AI.

Last but not least, our research deepens understanding of how paradoxical tensions in coopetition can be managed (Dagnino & Ritala, 2025). Platform owners can be conceptualized as a form of platform-level top management team (TMT) that mitigates competitive pressures and fosters cooperation by designing governance mechanisms for data access (Bengtsson & Raza-Ullah, 2025). In doing so, our framework complements strategy research emphasizing that openness is not entirely free. Instead, it generates the greatest benefits when pursued in a guided and targeted manner (Felin & Zenger, 2020). To this end, our study suggests several strategic approaches that platform owners can adopt to alleviate tensions.

Future Research Directions

Our framework points to several promising directions for future research, offering opportunities to advance understanding of the paradox of data openness and its implications.

First, while our theoretical framework provides the conceptual scaffolding to theorize the implications of data access decisions and the temporality of data for participant value creation and capture, it does not theorize the precise mechanisms that underlie these anticipated outcomes. Future research should seek to advance and empirically test research models that uncover what impact these data-related decisions have on participants and how those effects unfold in more specific contexts. For example, given that data is often a shared resource generated by multiple parties on a platform, one way to understand the cooptative tensions in value creation and capture is to investigate the ownership of digital assets. Blockchain technologies provide a potential remedy by providing immutable, transparent and verifiable records of data provenance and transactions (Lumineau et al., 2021; Lumineau et al., 2025). Through smart contracts, blockchain can codify rules for clarifying the ownership and usage rights of data assets. Future studies could explore, in the context of blockchain-enabled platforms, how rules governing data ownership alleviate or intensify tensions in value creation and capture among different actors within a digital platform ecosystem.

Second, future research can further investigate how tensions around data openness manifest at multiple levels (Ranganathan & Chen, 2025). Our study suggests that greater openness on one platform may inadvertently enhance the attractiveness of competing platforms. For example, rival platforms could leverage openly available data to train their own LLMs and improve services, thereby drawing users and complementors away. As such, tensions in value creation and capture may unfold not only between actors within a single platform, but also between multiple platforms.

Third, future research could adopt the perspective of complementors to examine how they engage in value creation and capture activities with varying degrees of data openness. While our theoretical framework describes how complementors may be affected by decisions regarding data access and the temporal nature of data, it does not provide insight into how they are likely to behave in the face of such situations. For example, when a platform imposes more restrictive data policies, complementors may respond by migrating to a more open platform or by multi-homing across several platforms. Research would benefit from efforts to theorize and empirically examine complementor decision making and competitive behavior within the context of digital platform ecosystems that are subject to the data-related decisions our framework articulates. For instance, it would be useful to learn how complementors incorporate these considerations into their decisions regarding the digital platforms on which to participate. It would also be useful to understand which complementor-related attributes might shape the impact of the framework's data-related decisions. In short, a theoretical and empirical understanding of complementors' behavioral responses is critical for advancing knowledge on value creation and capture in data-driven platform ecosystems.

Fourth, future research should explore how coopetition for data assets intersects with established strategy theories, such as dynamic capabilities, resource-based view, transaction cost economics, and stakeholder governance. For example, examining data openness through the lens of dynamic capabilities can illuminate how firms sense and seize opportunities and reconfigure their competence in rapidly evolving digital ecosystems. A resource-based view may highlight how data as a strategic asset underpins sustained competitive advantages and shapes firms' positioning in platform ecosystems. Transaction cost economics may shed light on how different governance arrangements for data access affect relationship-specific investments, value exchanges and opportunism. Stakeholder governance perspective may provide lens through which future research investigates the governance form that platform owners should take to navigate these tensions. In sum, exploring data openness tensions through these theoretical perspectives would enrich understanding of data-related coopetition and extend core strategy theories into the domain of open innovation on digital platforms.

Finally, we suggest future studies identify optimal solutions for resolving complex tensions (i.e., a combination of multiple tensions) in platform ecosystems. Notably, we do not presume that platform owners manage only one type of data (e.g., asymptotic data) or adhere to a single data openness strategy (e.g., open access to data). In reality, platform owners often have both asymptotic data and non-asymptotic data. A critical challenge, therefore, becomes how platform owners should configure their strategies. This may include balancing open and restricted access to data and navigating trade-offs between value creation and value capture for participants. Developing such optimal configurations becomes a challenge not only for platform owners but also a worthwhile research question.

Conclusion

New attention to the value of data has invited a reconsideration of key pillars of open collaboration. Data, and the intelligent algorithms that are trained on them, have become a prominent consideration for digital platforms as they pursue the value creation and value capture that sustains their existence. This new focus on the value of data has platform owners thinking through decisions as to whether to restrict access to the data that are generated from participant activity on their digital platforms or whether to open access to other parties. In this research, we addressed this evolving landscape by developing a framework that considers the paradox of data openness. We surface key tensions that arise in light of data-related decisions, both at the level of participants and at the level of the digital platform ecosystem. This offers one lens for strategic management scholarship to make sense of developments that will unfold regarding how platforms integrate data and proprietary algorithm-enabled services into their value proposition.

References

Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal*, *31*(3), 306-333.

Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction machines: The simple economics of artificial intelligence. In. Boston, MA: Harvard Business Review.

Alaimo, C., & Kallinikos, J. (2022). Organizations decentered: Data objects, technology and knowledge. *Organization Science*, *33*(1), 19-37.

Alexy, O., West, J., Klapper, H., & Reitzig, M. (2018). Surrendering control to gain advantage: Reconciling openness and the resource-based view of the firm. *Strategic Management Journal*, *39*(6), 1704-1727.

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., & von Hippel, E. (2011). Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization Science*, *22*(6), 1399-1417.

Baskerville, R. L., Myers, M. D., & Yoo, Y. (2020). Digital first: The ontological reversal and new challenges for information systems research. *MIS Quarterly*, *44*(2), 509-523.

Bengtsson, M., & Raza-Ullah, T. 2025. Paradoxical tensions at multiple levels and top management team cross-level bridging in coopetition: A conceptual model. *Strategic Management Review*, 6(1-2): 107-133.

Bercovitz, J., & Chesbrough, H. (2020). Hopping tables—An introduction to the SMR special issue on open innovation. *Strategic Management Review*, 1(2), 207-222.

Boudreau, K. (2010). Open platform strategies and innovation: Granting access vs. devolving control. *Management Science*, *56*(10), 1849-1872.

Boudreau, K. J., & Jeppesen, L. B. (2015). Unpaid crowd complementors: The platform network effect mirage. *Strategic Management Journal*, *36*(12), 1761-1777.

Burford, N., Shipilov, A. V., & Furr, N. R. (2022). How ecosystem structure affects firm performance in response to a negative shock to interdependencies. *Strategic Management Journal*, *43*(1), 30-57.

Cassiman, B., & Veugelers, R. (2006). In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition. *Management Science*, *52*(1), 68-82.

Ceccagnoli, M., Forman, C., Huang, P., & Wu, D. (2012). Cocreation of value in a platform ecosystem! The case of enterprise software. *MIS Quarterly*, *36*(1), 263-290.

Cennamo, C., & Santaló, J. (2013). Platform competition: Strategic trade-offs in platform markets. *Strategic Management Journal*, *34*(11), 1331-1350.

Chen, H., Chiang R., & Storey, V. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165-1188.

Chen, L., Mislove, A., & Wilson, C. (2016). An empirical analysis of algorithmic pricing on Amazon marketplace. *Proceedings of the 25th International Conference on World Wide Web*.

Chen, L., Tong, T. W., Tang, S., & Han, N. (2022). Governance and design of digital platforms: A review and future research directions on a meta-organization. *Journal of Management*, *48*(1), 147-184.

Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

Chesbrough, H. W., & Tucci, C. L. (2020). The interplay between open innovation and lean startup, or, why large companies are not large versions of startups. *Strategic Management Review*, 1(2), 277-303.

Chiambaretto, P., Fernandez, A.-S., & Le Roy, F. 2025. What coopetition is and what it is not: Defining the “hard core” and the “protective belt” of coopetition. *Strategic Management Review*, 6(1-2): 17-52.

Choudhury, V. (1997). Strategic choices in the development of interorganizational information systems. *Information Systems Research*, *8*(1), 1-24.

Clough, D. R., & Wu, A. (2022). Artificial intelligence, data-driven learning, and the decentralized structure of platform ecosystems. *Academy of Management Review*, *47*(1), 184-189.

CNBC. (2023). Twitter will start charging developers for API access as Elon Musk seeks to drive revenue. Retrieved 06/02/2023, from https://www.cnbc.com/2023/02/02/twitter-tostart-charging-developers-for-api-access.html

Constantinides, P., Henfridsson, O., & Parker, G. G. (2018). Introduction—Platforms and infrastructures in the digital age. *Information Systems Research*, *29*(2), 381-400.

Currier, J. (2020). What makes data valuable: The truth about data network effects. Retrieved 09/07/2023, from https://www.nfx.com/post/truth-about-data-network-effects

Dagnino, G., & Ritala, P. 2025. Coopetition strategy: Big questions and promising answers. *Strategic Management Review*, *6*(1-2), 1-5.

Dave, P. (2023). Stack Overflow will charge AI giants for training data. Retrieved 05/30/2023, from https://www.wired.com/story/stack-overflow-will-charge-ai-giants-for-trainingdata/

Eisenmann, T. R., Parker, G., & van Alstyne, M. (2009). Opening platforms: How, when and why. *Platforms, Markets and Innovation*, *6*, 131-162.

Engert, M., Hein, A., Maruping, L. M., Thatcher, J. B., & Krcmar, H. (2025). Self-organization and governance in digital platform ecosystems: An information ecology approach. *MIS Quarterly*, 49(1), 91-122.

Enkel, E., Gassmann, O., & Chesbrough, H. (2009). Open R&D and open innovation: Exploring the phenomenon. *R&D Management*, *39*(4), 311-316.

Farrell, J., & Simcoe, T. (2012). Four paths to compatibility. In *The Oxford Handbook of the Digital Economy* (pp. 34-58). Oxford University Press Oxford, UK, and New York.

Felin, T., & Zenger, T. R. (2020). Open innovation: A theory-based view. *Strategic Management Review*, 1(2), 223-232.

Garud, R., Jain, S., & Kumaraswamy, A. (2002). Institutional entrepreneurship in the sponsorship of common technological standards: The case of Sun Microsystems and Java. *Academy of Management Journal*, *45*(1), 196-214.

Gawer, A., & Cusumano, M. A. (2008). How companies become platform leaders. *MIT Sloan Management Review*, *49*, 28.

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.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The role of artificial intelligence and data network effects for creating user value. *Academy of Management Review*, *46*(3), 534-551.

Hausman, J. A. (1981). Exact consumer's surplus and deadweight loss. *The American Economic Review*, *71*(4), 662-676.

Hawkins, A. J. (2023). Google Maps is becoming more like search — Thanks to AI. Retrieved 03/30/2024, from https://www.theverge.com/2023/10/26/23932315/google-maps-aiimmersive-view-ev-charging-search

Huang, J., Henfridsson, O., & Liu, M. J. (2022). Extending digital ventures through templating. *Information Systems Research*, *33*(1), 285-310.

Huang, S., & Siddarth, D. (2023). Generative AI and the digital commons. *arXiv preprint arXiv:2303.11074*.

Hukal, P., Henfridsson, O., Shaikh, M., & Parker, G. (2020). Platform signaling for generating platform content. *MIS Quarterly*, *44*(3), 1177-1205.

Issac, M. (2023). Reddit wants to get paid for helping to teach big A.I. Retrieved 05/30/2023, from https://www.nytimes.com/2023/04/18/technology/reddit-ai-openai-google.html

Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, *39*(8), 2255-2276.

Karhu, K., Gustafsson, R., & Lyytinen, K. (2018). Exploiting and defending open digital platforms with boundary resources: Android's five platform forks. *Information Systems Research*, *29*(2), 479-497.

Krämer, J., & Shekhar, S. (2025). Regulating digital platform ecosystems through data sharing and data siloing: Consequences for innovation and welfare. *MIS Quarterly*, *49*(1), 123-154.

Kretschmer, T., Leiponen, A., Schilling, M., & Vasudeva, G. (2022). Platform ecosystems as meta-organizations: Implications for platform strategies. *Strategic Management Journal*, *43*(3), 405-424.

Laursen, K., & Salter, A. J. (2014). The paradox of openness: Appropriability, external search and collaboration. *Research Policy*, *43*(5), 867-878.

Laursen, K., & Salter, A. (2020). Who captures value from open innovation—the firm or its employees? *Strategic Management Review*, 1(2), 255-276.

Lei, X., Chen, Y., & Sen, A. (2023). The value of external data for digital platforms: Evidence from a field experiment on search suggestions. *Available at SSRN*.

Leswing, K. (2022). Facebook says Apple iOS privacy change will result in $10 billion revenue hit this year. Retrieved 09/06/2023, from https://www.cnbc.com/2022/02/02/facebooksays-apple-ios-privacy-change-will-cost-10-billion-this-year.html

Levine, S. S., & Prietula, M. J. (2014). Open collaboration for innovation: Principles and performance. *Organization Science*, *25*(5), 1414-1433.

Lin, Y. K., & Maruping, L. M. (2022). Open source collaboration in digital entrepreneurship. *Organization Science*, *33*(1), 212-230.

Liu, S., Henfridsson, O., Hummel, J. T., & Nandakumar, J. (2024). The complementor's dilemma: Navigating growth ambitions and the dependency on focal actors in platform ecosystems. *MIS Quarterly*. Forthcoming.

Loh, J., & Kretschmer, T. (2023). Online communities on competing platforms: Evidence from game wikis. *Strategic Management Journal*, *44*(2), 441-476.

Lumineau, F., Shang, G., Swaminathan, J., Tsoukalas, G., Wagner, S., & Zhao, L. 2025. Charting the future of blockchain in operations and supply chain management: Opportunities and challenges. *Journal of Operations Management*, Forthcoming.

Lumineau F., Wang W., & Schilke O. 2021. Blockchain governance—A new way of organizing collaborations? *Organization Science, 32*(2), 500-521.

Minà, A., & Dagnino, G. B. (2025). Coopetition reloaded: Looking back to strategic management research for moving coopetition inquiry forward. *Strategic Management Review, 6*(1-2), 135-191.

McIntyre, D. P., & Srinivasan, A. (2017). Networks, platforms, and strategy: Emerging views and next steps. *Strategic Management Journal*, *38*(1), 141-160.

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management. *MIS Quarterly*, *41*(1), 223-238.

Niculescu, M. F., Wu, D., & Xu, L. (2018). Strategic intellectual property sharing: Competition on an open technology platform under network effects. *Information Systems Research*, *29*(2), 498-519.

O'Mahony, S., & Karp, R. (2022). From proprietary to collective governance: How do platform participation strategies evolve? *Strategic Management Journal*, *43*(3), 530-562.

Panico, C., & Cennamo, C. (2022). User preferences and strategic interactions in platform ecosystems. *Strategic Management Journal*, *43*(3), 507-529.

Parker, G., & van Alstyne, M. (2018). Innovation, openness, and platform control. *Management Science*, *64*(7), 3015-3032.

Parker, G., van Alstyne, M., & Choudary, S. (2016). How networked markets are transforming the economy and how to make them to work for you. In: New York: W.W. Norton Publishing.

Parker, G., van Alstyne, M. W., & Jiang, X. (2017). Platform ecosystems: How developers invert the firm. *MIS Quarterly*, *41*(1), 255-266.

Peters, J. (2023). Apollo for Reddit is shutting down. Retrieved 06/13/2023, from https://www.theverge.com/2023/6/8/23754183/apollo-reddit-app-shutting-down-api

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192-210.

Ranganathan, R., & Chen, J. (2025). The antecedents and consequences of multi-firm technology coordination: An ecosystem perspective. *Strategic Management Review*, *6*(1-2), 75-106.

Rietveld, J., Schilling, M. A., & Bellavitis, C. (2019). Platform strategy: Managing ecosystem value through selective promotion of complements. *Organization Science*, *30*(6), 1232-1251.

Roth, E. (2023). Microsoft, GitHub, and OpenAI ask court to throw out AI copyright lawsuit. Retrieved 09/07/2023, from https://www.theverge.com/2023/1/28/23575919/microsoftopenai-github-dismiss-copilot-ai-copyright-lawsuit

Ryan-Charleton, T., & Gnyawali, D. R. (2025). Value Creation Tension in Coopetition: Virtuous and Vicious Cycles. *Strategic Management Review*, *6*(1-2), 193-218.

Say, G., & Vasudeva, G. (2020). Learning from digital failures? The effectiveness of firms' divestiture and management turnover responses to data breaches. *Strategy Science*, *5*(2), 117-142.

Schreieck, M., Wiesche, M., & Krcmar, H. (2016). Design and governance of platform ecosystems – Key concepts and issues for future research. *European Conference on Information Systems.*

Seidel, V. P., Langner, B., & Sims, J. (2017). Dominant communities and dominant designs: Community-based innovation in the context of the technology life cycle. *Strategic Organization*, *15*(2), 220-241.

Shah, S., & Nagle, F. (2019). Why do user communities matter for strategy? *Strategic Management Review*, *1*(2), 305-353.

Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review*, *36*(2), 381-403.

Teece, D. J. (2020). Hand in glove: Open innovation and the dynamic capabilities framework. *Strategic Management Review*, *1*(2), 233-253.

Tian, J., Zhao, X., & Xue, L. (2022). Platform compatibility and developer multihoming: A tradeoff perspective. *MIS Quarterly*, *46*(3), 1661-1690.

Tilson, D., Lyytinen, K., & Sorensen, C. (2010). Digital infrastructures: The missing IS research agenda. *Information Systems Research*, *21*(4), 748-759.

Tong, A., Wang, E., & Coulter, M. (2024). Exclusive: Reddit in AI content licensing deal with Google. Retrieved 03/20/2024, from https://www.reuters.com/technology/reddit-aicontent-licensing-deal-with-google-sources-say-2024-02-22/

van der Vlist, F., Helmond, A., & Ferrari, F. (2024) Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence. *Big Data & Society*, *11*(1): 1–16.

Vasudeva, G., Leiponen, A., & Jones, S. (2020). Dear enemy: The dynamics of conflict and cooperation in open innovation ecosystems. *Strategic Management Review*, *1*(2), 355-379.

von Hippel, E., & von Krogh, G. (2003). Open source software and the "private-collective" innovation model: Issues for organization science. *Organization Science*, *14*(2), 209-223.

Wormald, A., Shah, S. K., Braguinsky, S., & Agarwal, R. (2023). Pioneering digital platform ecosystems: The role of aligned capabilities and motives in shaping key choices and performance outcomes. *Strategic Management Journal*, *44*(7), 1653-1697.

Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564*.

Xue, L., Song, P., Rai, A., Zhang, C., & Zhao, X. (2019). Implications of application programming interfaces for third−party new app development and copycatting. *Production and Operations Management*, *28*(8), 1887-1902.

Zhang, Y., Li, J., & Tong, T. W. (2022). Platform governance matters: How platform gatekeeping affects knowledge sharing among complementors. *Strategic Management Journal*, *43*(3), 599-626.

Zhu, K., Kraemer, K. L., Gurbaxani, V., & Xu, S. X. (2006). Migration to open-standard interorganizational systems: Network effects, switching costs, and path dependency. *MIS Quarterly*, *30*, 515-539.

Zuboff, S. (2019). Surveillance capitalism and the challenge of collective action. In *New Labor Forum* (Vol. 28, pp. 10-29). SAGE Publications Sage CA: Los Angeles, CA.