Initiating and expanding data network effects: A longitudinal case study of generativity in the evolution of an AI platform

Maria Kandaurova
Chalmers University of Technology
maria.kandaurova@chalmers.se

Daniel A. Skog
Umeå University
daniel.a.skog@umu.se

Abstract

This study explores the emergence and expansion of data network effects (DNEs) in AI platforms. Previous research has focused on direct and indirect network effects. However, the rise of AI platforms necessitates understanding DNEs for platforms' learning and improvement. Through a longitudinal case study of a Conversational AI (CAI) platform's 12-year evolution, the study identifies generative feedback loops as the mechanism for DNEs. These loops are initiated by adding functions that enhance the platform's generative capacity, resulting in more diverse data that improves platform learning. DNEs develop through interactions with different ecosystem actors, including clients and external developers, and rely on various data sources beyond user data to enhance AI platform capabilities. This study contributes to IS literature, specifically digital platform literature, following recent calls to empirically examine DNEs to better understand how AI platforms grow and improve their algorithmic capabilities over time.

Keywords: Data Network Effects, Generativity, Digital Platforms, AI Platforms, Longitudinal Case Study

1. Introduction

AI platforms are digital platforms whose value creation and capture centers on enabling organizations to apply AI in the augmentation and automation of tasks. As a particular type of digital platform, AI platforms develop through the initiation of network effects, i.e., when growth in the ecosystem makes the platform more valuable and attracts more actors to the ecosystem in a form of self-reinforcing mechanism (Parker, Van Alstyne, & Jiang, 2017).

The proliferation of AI platforms that rely on large amounts of data to improve algorithmic capabilities have highlighted the initiation, sustainment and expansion of Data Network Effects (DNEs) as key (Gregory, Henfridsson, Kaganer, et al., 2021). In the context of AI platforms, DNEs are defined as the phenomenon by which an increase in user generated data makes a platform more valuable to each user. DNEs are inherently dependent on the platform's AI capability, i.e., its ability "*to learn from data to continuously improve its products or services*" (Gregory, et al., 2021, p. 535). To this point, however, empirical studies of DNEs are lacking, and as a result, we have limited knowledge about whether and how they emerge during the evolution of AI platforms.

Against this backdrop, the aim of this paper is to empirically explore the novel phenomenon of DNEs and their role in AI platform evolution. To that end, the paper is guided by the following research question: *how do data network effects emerge and expand in AI platforms over time?*

We explore this through a case study of the evolution of the CAI-platform, a conversational AI (CAI) platform provided by CAI-Corp. Started in the early 2000s, CAI-Corp has been providing AI-simulated human conversation software products for commercial purposes long before the introduction of Apple's Siri and ChatGPT. Our analysis starts in 2010 when CAI-Corp fundamentally alters its offering from specific applications for particular tasks and clients to the modular CAI-platform enabling the exchange of resources between an expanding set of actors in a wider ecosystem.

By tracing the evolution of the CAI-platform to the present day, we identify episodes of DNEs initiation and expansion. These episodes involved making specific platform enhancements to cater to the needs of either clients or developers. Beyond satisfying their direct needs, each enhancement also enabled these actors to take new forms of actions that extended the scope and scale of data available for the platform to learn from. In turn, these insights were often leveraged for making further platform enhancements which enabled new forms of actions amongst ecosystem actors.

Drawing on the concept of generativity (Thomas & Tee, 2022), we explain how episodes represent generative feedback mechanisms, where intermittent enhancements to the platform's generative capacity enable outcomes that ultimately come to shape the platform, often resulting in further expansion to its generative capacity. We theorize how DNEs may

URI: https://hdl.handle.net/10125/107135
978-0-9981331-7-1

emerge and expand in AI platforms over time through such generative feedback mechanisms and detail three distinct contributions to the literature.

2. Theoretical background

2.1 Network effects as growth mechanisms in platform ecosystems

AI platforms are digital platforms that leverage AI technologies as an inherent part of their core service (Rai, Constantinides & Sarker, 2019; Sundberg & Holmström, 2023). As all digital platforms, AI platforms are essentially digital products or services that mediate resource transactions between producers and users (de Reuver, Sørensen, & Basole, 2018). Rather than relying solely on internal resources and capabilities for value creation, digital platforms provide interfaces through which external producers can provide complementary resources (Gawer, 2021). Catering to the user side, digital platforms have interfaces through which users can access and use these resources, which in the case of AI platforms often manifests in low-code or no-code graphical user interfaces (GUI) (Sundberg & Holmström, 2023).

The network of actors that use a digital platform for resource transactions are conceptualized as platform ecosystems (de Reuver et al., 2018). Notwithstanding the tendency in research to focus on app developers, platform ecosystems may consist of a range of diverse actors that complement the core service of the platform in different ways. For AI platforms, providers of data, content, functions, or machine learning (ML) models are likely to be as important as app developers. Considering the capacity of AI platforms to learn and improve their capabilities over time, end-users, and the data they transact through the platform over time also represent key sources of ecosystem value.

The growth of digital platforms and ecosystems is mutually dependent, driven by a form of self-reinforcing mechanism known as network effects. In essence, network effects manifest when an increase in actors using the platform also extends the value of the platform, which in turn attracts more actors to it (Parker et al., 2017). Traditionally, the literature has focused on two categories of the network effects that digital platforms exhibit. First, direct network effects refer to when the platform's value increases for one type of actor as more actors of the same category join the platform, such as when friends join a social media platform. Second, indirect network effects arise when growth in one category of ecosystem actors adds value to the platform that attracts another type of actor, for example when a growth in app developers extends service variety that attracts more end-users. The attainment of both effects relies on the same factors, namely the ability of the platform to attract different forms of ecosystem actors, to ensure that the resources they provide to the platform are of a desired quality, and to make resource transactions increasingly easy, beneficial and valuable to all ecosystem actors.

This is primarily achieved through platform owners' careful adjustment of technical interfaces and governance systems (Eaton, Elaluf-Calderwood, Sørensen, et al., 2015; Skog, Wimelius, & Sandberg, 2018). At the same time, recent research also suggests that different types of network effects can play different roles based on what core service a digital platform offers (Skog, Sandberg, & Wimelius, 2021). For AI platforms in particular, their ability to both enable data transactions at a massive scale and to learn and improve based on these transactions, motivates deeper examination of potentially novel forms of network effects (Gregory, et al., 2021). Recognizing these unique characteristics, Gregory et al. (2021) argue that we need to consider data network effects as a new type of network effect to extend our understanding of the utility, evolution, and competitiveness of AI platforms. Similarly, to other forms of network effects, DNEs rest on attaining sufficient data transactions, both in terms of quantity and quality, and on the ability of the platform to leverage these transactions for learning and improving its service (Gregory, et al., 2021).

However, despite the apparently key role of establishing and maintaining network effects, there is a general lack of knowledge of longitudinal platform dynamics (de Reuver et al., 2018) that could help us understand how this can be achieved. For AI platforms in particular, very few empirical studies have been done on whether, what and how network effects emerge and expand over time. From the literature on digital platforms, we may assume these effects to emerge at the intersection between digital platform capacities and actors within the surrounding ecosystem, a relationship that can be more deeply understood through the concept of generativity.

2.2 Generativity

The concept of generativity, broadly understood as a system's capacity to create or produce new and innovative outcomes beyond its original design, has become increasingly used to explain certain forms of open digital innovation dynamics (Thomas & Tee, 2022), particularly those that characterize and drive the evolution of digital platforms (Sun, Xu, & Karanasios, 2023). In the digital platform literature, generativity has been used in relation to how digital platform providers can attract and leverage third-party developers to extend the value of their core product or service with complementary service applications (Ghazawneh & Henfridsson, 2013; Parker et al., 2017).

The concept of generativity has predominantly been seen as either a capacity of technology (Fürstenau, Baiyere, Schewina, Schulte-Althoff, & Rothe, 2023; Henfridsson & Yoo, 2014; Um, Yoo, Wattal, Kulathinal, & Zhang, 2013) or an outcome of interactions between human actors and technologies (Bygstad, 2017; Staub, Haki, Aier, & Winter, 2022), yet more recent work proposes a view of generativity as a sociotechnical system (Thomas & Tee, 2022).

The above views can be traced to Zittrain's original work (2008) where generativity is described both in terms of generative technologies with certain capacities or as something that is denotative (or a function) of these capacities (hence not a technological capacity but something that emerges from use). In terms of technological capacities, Zittrain (2008) proposed that generative technologies must be easily accessible, adoptable, leveraged, and adaptable for a wide range of users, while also being highly efficient and effective for accomplishing diverse tasks. Accessibility relates to the ease with which a wide range of users can access the technology. Adoptability refers to how easily it can be used and reconfigured, allowing for broad adoption by different audiences. The technology's leveraging capacity refers to its ability to handle diverse tasks. Adaptability refers to how easily the system can be built upon or modified, expanding its range of applications. These generative capacities have been argued to be prominent in digital artifacts (Eck, Uebernickel, & Brenner, 2015), and in digital platforms in particular. To that end, an inherent capacity to be reprogrammed, modular, and loosely-coupled have all been argued to further improve the generative capacity of digital platforms (Yoo et al., 2010).

In this paper, we build on the sociotechnical perspective and hence we direct our focus to interactions between the generative capacity of technology and the user community that utilizes it as the locus of generativity. From this perspective, generativity manifests when the generative capacity of platforms is leveraged by a user community to do new things that ultimately results in new improvements to the platform, here referred to as a generative feedback loop (Thomas & Tee, 2022). To enable generative feedback loops, providers may increase the generative capacity of platforms by ensuring that the technical architecture offers a sufficient degree of malleability, granularity and loose-coupling, but also by ensuring that governance structures are tuned to facilitate platform access and use. Ultimately, it is through the ability of the user community to access and use platform resources for new actions that generative feedback loops can be initiated and maintained.

At the core of the concept of generativity lies the notion of balancing tensions as necessary to gaining value from the potential variety of resources generated by a user community. To initiate and harness generativity, platform owners must adhere to the Goldilocks principle of finding the right balance - neither too little nor too much, but just right. This entails maintaining a stable technical infrastructure that facilitates connections with ecosystem actors, yet remains flexible enough for expansion (Tilson, Lyytinen & Sørensen, 2010). Similar considerations apply to governance mechanisms, which must strike a balance between providing access and autonomy to external actors to foster variety, while ensuring only the desired level of variety is incorporated back into the platform (Wareham, Fox, & Giner, 2014).

Digital platforms may cater to and benefit from different types of users over time (Wareham et al., 2014; Skog et al., 2021). For AI platforms in particular, we may assume that all ecosystem actors whose use of the platform may extend the scope and scale of digital data or its AI capability are potentially relevant to the initiation and maintenance of DNEs.

3. Methodology

We employed a longitudinal case study design to gain insights into how data network effects (DNEs) emerge and expand in AI platforms over time with the intent to make analytical generalizations about this process (Yin, 2018). To that end, we investigated the development of a CAI platform from 2010 to 2022.

Current IS literature defines CAI as *"a general capability of computers to understand and respond with natural human language as it is written or spoken."* (Benbya et al., 2021, p. 302). To gain this capability, platforms may combine different AI technologies including natural language processing (NLP), machine learning (ML), automatic speech recognition (ASR) and speech-to-text (STT). The case CAI-platform utilizes such technologies to enable the development of conversational agents (CAs), i.e., bots that communicate with humans through text and voice.

We selected the CAI-platform as a case for three main reasons. First, its relatively long history as an AI platform provided us with the opportunity to study the longitudinal dynamics of DNEs. Second, like any other CAI platform it depends on a substantial amount of data to handle intricate tasks like language input, processing, and output, enabling conversational agents to deliver more human-like conversations. Consequently, the development of CAI platforms is inherently dependent on DNEs. Lastly, the AI capability of the CAI-platform combines both ML and rule-based algorithms, and hence enables us to study the generation of different types of data. While the ML component relies on larger datasets to learn and improve over time, the rule-based component depends on smaller data sets that can enable faster deployment and reduce the barriers for users to engage with the system.

3.1. Data collection

We collected data from multiple sources to conduct our longitudinal case study (Yin, 2018), including both secondary and primary data. First, we collected 209 blog posts and 294 press releases published on CAI-Corp's website between 2010 and 2022. These posts covered various topics including industry challenges, changes to the platform's architecture, alterations to developer resources, new market entries, partnership agreements, use cases, and other key development events. We also obtained 16 datasheets (i.e., documents summarizing technology's components, specifications, and characteristics) and two presentations that summarized the platform and its respective modules. Additionally, we utilized the Internet Archive Wayback Machine to accurately deduce the temporal sequence of particular incidents on CAI-Corp's website at specific time intervals. Together, these sources allowed us to outline key events in the evolution of the CAI-platform in chronological order.

Finally, we conducted six interviews with CAI-Corp NLI/AI experts and business managers, referred to as case company informants (CCIs), to understand "*why and how certain decisions or choices* [were] *made, at specific moments in time"* (Cloutier & Langley, 2020, p.5). Each interview lasted for 60 minutes, and was transcribed using otter.ai.

3.2. Data analysis

In analyzing our data, we followed process analysis methods suited to understanding the how and why of evolving phenomena (Langley, 1999). Albeit highly iterative in practice, our analysis approach can be described in three linear steps. First, we developed a chronological timeline of key events in the evolution of the CAI-platform from 2010. To determine the relevance of events in our collected data, we drew on Skog et al. (2018) and included those we deemed related to the key dimensions in digital platform evolution, i.e., architecture, governance, and ecosystem (Tiwana, Konsynski, & Bush, 2010). This resulted in the exclusion of data related to e.g., competitions, awards and marketing campaigns, while data points regarding new module launches, resource announcements, organizational strategies, rules and regulations, new clients, use cases, and partnerships were included. This process resulted in a reduction from 509 to 165 relevant data points used for the development of the timeline. Using the Aeon Timeline software, we organized these data points into discrete chronological events on a visual timeline, each consisting of a title, summary, date of occurrence and a verbatim description.

Second, we analyzed timeline events in detail to identify possible indications of DNEs, i.e., instances where the CAI-platform had become more valuable to its users due to data generated by use. By identifying major changes to the platform's architecture and tracing these back and forward in time, we noticed patterns of interaction between an architectural change, new actions amongst a particular type of ecosystem actor, and new architectural changes. We treated these identified patterns as analytical episodes, i.e., distinct sequences of events of potential relevance to what we aimed to theorize.

Finally, seeing that it could help us understand and explain what was going on in the identified episodes, we drew on the literature on generativity (Thomas & Tee, 2022). From the perspective on generativity as the result of interactions within a sociotechnical system, we analyzed events of architectural changes as potential enhancements to the generative capacity of the platform. If and when these architectural changes led to new forms of actions amongst ecosystem actors, and data generated from these new actions was evidently used to make subsequent improvements to the platform, we conceptualized them as generative feedback loops.

4. Findings

In what follows, we first describe events that, while predating the period of our analytical focus, explain and provide important preconditions for later actions taken to initiate and extend data network effects. Second, we present parallel processual trajectories identified in our analysis through which distinct forms of data network effects were developed in relationship to different ecosystem actors. Within each trajectory, we detail chronological episodes describing how DNEs were enabled through generative feedback mechanisms.

4.1. Prologue

CAI-Corp originated in the early 2000s as a customer service optimization (CSO) software suite provider. At the core of its CSO software were rule-based algorithms that enabled the creation of custom CAs that could autonomously process and respond to customer enquiries. In essence, this AI capability consisted of hand-coded rules determining correct responses to enquiries in an "if-this-then-that" fashion. CAI-Corp relied heavily on computational linguists to leverage rule-based natural language interaction (NLI) to create CAs that were adaptive towards diverse business domains and languages. However, manually improving and maintaining multiple custom CAs implemented in different client contexts was labor-intensive and costly, and the CAI-Corp therefore sought ways to manage them automatically at scale. Around 2009, the increasing use of smartphones and demands for having multi-channel customer interactions further highlighted the need for more cost-effective and scalable solutions.

In response, CAI-Corp embarked on developing what they referred to as a *"managerial and scalable platform"* in 2010. This entailed transforming the suite of CSO software applications into a comprehensive and modular CAI platform. By 2013, the CAI-platform consisted of three loosely coupled modules: a module for the NLI capability which previously served as a backbone of the CSO suite, an analytics module, and a CA customization module. The purpose of these native modules was to establish and enhance the AI capability of the platform over time, while focusing on advancing NLI, getting access to data transactions in both quantity and quality, and providing a GUI tailored to non-technical users. The transition from a software suite to a modular platform marked a shift in CAI-Corp's approach from predominantly in-house development to a more distributed, collaborative, and data-driven approach. Following this shift, CAI-Corp increasingly relied on its broader ecosystem to capture and leverage data, consequently enhancing its AI capabilities. Below, we trace the evolution of the platform between 2010-2022 through two parallel trajectories focused on the relationship between the platform, clients, and developers.

4.2. Client trajectory

Episode 1: Enabling core conditions for platform-client data network effects (2010-2015)

CAI-Corp's primary client group consisted of companies that used the platform to create chat or voice bots to provide customer or employee support. Our analysis showed that CAI-Corp's strategic focus to meet clients' needs resulted in the gradual development and improvement of its capacity to collect, analyze, and improve AI capabilities through data. This progress has been accomplished by establishing generative feedback mechanisms that foster interaction and collaboration between the platform and the broader ecosystem.

A notable step to improve CAI-Corp's data capability was the introduction of an analytics module in early 2011. Initially, CAI-Corp purchased the analytics module as a plug-in from an external Business Intelligence (BI) company. While this integration extended the platform's functionalities for clients, it was limited to the BI's provider who offered limited flexibility in terms of when and in what format clients could analyze data. More importantly, CAI-Corp was unable to access the data used within the analytics module as it was stored in the BI provider's infrastructure rather than its own.

To overcome this limitation, CAI-Corp developed a proprietary analytics module in late 2011**.** This provided two main benefits. First, customizing the analytics module towards its own platform allowed CAI-Corp to implement more advanced functionalities, including a more accessible and highly scalable data repository, improved data management, and various interfaces that allowed clients to export datasets to external BI tools like Power BI or Tableau. Second, it allowed CAI-Corp to access all the "*conversational data and metadata, whether when having a voice or text interaction, whatever it is, to have all that data in one place*" (CCI4). These adaptations helped gain new clients and collect and analyze more usage data.

A second notable key improvement to the platform architecture was the introduction of a CA customization module in 2012. This module provided clients with an intuitive, visual, and user-friendly tool to create CAs and customize their interactions with customers. Its low-code nature made it accessible to non-technical users, including business unit managers and customer service agents. Notably, this module not only expanded the *accessibility* of CA development to a broader user base but also made the platform more *adaptable* to a wide range of domain and country-specific requirements. Consequently, the platform became increasingly attractive to large enterprises seeking global scalability while reducing their dependence on technical resources. Additionally, this module allowed CAI-Corp to access a diverse range of conversational data, providing insights into domain-specific multilingual conversations. As expressed by our informants, it was key to "*sit on the correct type of data* […] *the closer your data is to the domain, the better*" (CCI1). This access to diverse data allowed CAI-Corp to expand its language models to support more languages, further enhancing its ability to serve clients operating in various linguistic contexts.

However, although the platform came to excel in accurately simulating NLI-based conversations, improvements to this capability relied heavily on organic, iterative improvements backed by rule-based algorithms, which limited the ability of CAs to effectively handle unexpected data patterns (e.g., misspelled words, a sudden change of subject, etc.) As put forth by our informants, *"the NLI is quite expressive. People make mistakes when they communicate, they misspell words, change subjects, and it's hard to account for all these nuances."* (CCI2).

CAI-Corp and the market environment around 2015 saw multiple changes. First, AI capabilities based on rule-based NLI *"started to sink"* and CAI-Corp *"was looking for ways to modernize the platform"* (CCI1). Competitors such as IBM, Microsoft, Google *"started coming up with all these ML-based platforms* [for developing CAs]. *First, we had API.AI that became Google Dialogflow, then Amazon came up with Lex, and Microsoft with Luis. We had been in this space for a long time. And we knew our biggest competitors back then. But suddenly, our competitors were Google, Amazon and all the big technology companies"* (CCI1). Clients started to ask *"whether* [the platform] *offers machine-learning capabilities, as many organizations started to use it"* (CCI3). This moved CAI-Corp to leverage ML capabilities without losing the advantage of its advanced rule-based capabilities.

Episode 2: Towards a hybrid AI capability (2015-2022)

In 2015, CAI-Corp incorporated ML into the platform by leveraging open-source third-party technology. Informants explain how, since the platform was modular, *"it was quite straightforward to plug in a third-party tool to do the ML classification of the inputs…. So, the clients could early on start using ML."* (CCI4). Although the solution initially improved the platform's *"robustness by handling errors, variation in NLI, and large datasets while achieving higher recall* [i.e., system's ability to remember relevant information that might not be explicitly defined by the rule-based approach]" (CCE1), its long-term prospects were weak. The company clients were restricted to this one particular third-party technology. Moreover, *"training of models was managed externally to the platform"*, hindering the improvement of the platform's core capability over time. Informants further explain how, *"it would require additional tools or workflows outside of the platform, which would introduce additional complexity or inefficiencies in the training process."* (CCI4).

Albeit exploring the potential of ML, CAI-Corp did not abandon the capability in rule-based NLI it had developed to this point. This is because in contrast to ML, it allowed the platform and CAs to have *"the fine-grained understanding of the language nuances and deliver even higher precision than the one offered by ML alone,* [i.e., system's ability to consistently and accurately comprehend user intent and deliver the desired results]" (CCI1). Rather, as expressed by informants, it became apparent that ML and rule-based approaches could be combined to create a form of a hybrid AI capability: *"I had been pushing for the ML approach before, and it really helped me to convince the senior managers that it was the way to go, when the clients started asking about it. Though we could completely switch to ML, I did not believe in it fully. It had limitations. I was advocating for incorporating both approaches and offering a much stronger, hybrid architecture to our clients, which would be flexible, yet stable and enable them to have more control, which is good when you have enterprise clients"* (CCI1).

However, reflecting on prior experiences with the analytics module, CAI-Corp and its clients saw how relying solely on one third-party ML technology came to restrict the functional variety, accessibility and adaptability of the platform. For example, they had no control over, or insight into, the training of ML models, which restricted the platform from fully exploiting the synergy potential with the platform's established NLI core capability. To address this, CAI-Corp decided to "*embed the third-party tool within the product to do ML classification and train the data within the platform, rather than having to do it outside of it*" (CCI4). This native to the platform ML capability could be easily recombined with existing GUI development environment enabling the platform to "*offer some extra screens in the UI*" that enabled clients to "*see how balanced their ML models were and manage a bit more the ML classification*" (CCI4).

As a result, the integration of ML and the transition to a hybrid CAI platform, generated a feedback loop reminiscent of episode one. The addition of a native ML component increased the platform's capability to deal with and learn from large datasets captured by CAs and then reuse these data to improve the platform. Informants share how adding ML leveraged higher flexibility and scalability of the platform, stating *"with ML, platform's maintenance could be done simpler because of the data you collect. You can then just retrain the models based on that. At the same time, you have to have some control of that process. Because otherwise, it can go wildly wrong and make things worse. So, it's not a completely free process. There is no free lunch here. You need to find methods or processes to select good from bad data to integrate into the model"* (CCI1). Thus, shifting to a hybrid approach enabled better control over the CAI system's behavior. It combined the flexibility and learning capacity of ML with the reliability and control offered by the rule-based approach. Informants explain how this hybrid approach with its doubled capacity generated an increase in the scope and scale of data, and attracted more clients, because they could build *"much more complex systems that could pick up subtle patterns in the variety of NLI, while controlling for higher precision and predictability, and that was the advantage of our platform"*(CCI1).

The above episodes illustrate how in catering to the clients' needs, CAI-Corp became more accessible, adaptable, and leveraging for different tasks. Despite lacking self-learning AI capabilities in their rule-based algorithms at the time, CAI-Corp leveraged the data collected through the various adaptations of their platform to manually enhance the algorithms. This data proved to be instrumental in refining the performance and effectiveness of the algorithms, enabling CAI-Corp to make iterative improvements based on real-world usage and feedback. By utilizing the insights gained from the data, CAI-Corp was able to optimize their algorithms and enhance the overall AI capabilities of their platform, delivering more accurate and efficient results for their clients.

4.3. Developer trajectory

Episode 3: Enabling conditions for platform - developer data network effects (2011 - 2022)

With Apple's Siri entering the conversational AI landscape, *"the launch exerted significant pressure on the case company",* prompting them to adapt to the new landscape. However, the company had an advantage of being an early participant in the field, having existed in the *"pre-Siri"* era (CCI1). In 2011, CAI-Corp expanded its platform into the mobile ecosystem, enabling developers to quickly deploy their own mobile CAs across different devices and operating systems. Additionally, to showcase the platform's capabilities, the company introduced multilingual CAs capable of conversing in over 20 languages. This move facilitated a generative feedback mechanism with a broader community of developers and private users in the mobile ecosystem, providing access to *"diverse language-based data and improving the platform's language models and core capabilities"* (CCI4).

In 2012, CAI-Corp introduced a development environment that empowered developers consisting of a set of tools that enabled developers to make the capabilities of the platform and its new mobile version more accessible. This environment provided tools for designing, building, and testing CAs, along with APIs, web services, and code samples for seamless integration and customization. It allowed developers to explore potential use cases and unlock the platform's capabilities. By 2019, the latest version of the development environment offered advanced features such as analytics, multilingual support, and integration with various channels and third-party platforms. The platform's utilization was further amplified during the COVID-19 pandemic, as it assisted new partners and clients in developing CAs and expanded to include *"domain-specific business-focused CAs for remote HR and IT support"*. These developments attracted more clients and enriched the platform's data, thereby enhancing its AI capabilities.

In 2019, a leading technological research and consulting firm, emphasized the increasing significance of seamless integration between CAI platforms and other systems. Long-standing clients of CAI-Corp *"expressed their need for specific integrations with their back-end systems"* (CCI5), prompting the company to develop interfaces to ensure seamless functionality. This shift toward greater adaptability and accessibility characterized the platform's evolution between 2020 and 2022 because of a growing number of pre-build interfaces.

Episode 4: Exploiting the generative community of developers (2019 - 2022)

To meet client requirements and align with the broader ecosystem, CAI-Corp introduced two initiatives. Firstly, they launched a new component, a resource that included pre-built interfaces and plugins to facilitate interaction between the platform and external back-end systems such as robotic process automation (RPA). This expansion enabled CAs to provide intelligent responses by leveraging actions and processes from the back-end systems. For example, CAs could handle tasks like loan approvals or HR events based on employee leave requests. The platform's enhanced adaptability attracted more clients, resulting in increased data and further expansion of its scope and scale.

Secondly, CAI-Corp focused on expanding the library of pre-built interfaces and plugins by establishing closer ties with a diverse community of developers. CAI-Corp created a developer knowledge space to provide access to documentation, tutorials, and code samples, fostering collaboration and support within the community. Though CAI-Corp tried to build various interfaces in-house, in the early stages, *"it was more time critical"*. To accelerate development, the company engaged the developer community; *"those hardcore programmers who would hang out on Stack Overflow"*, through campaigns and monetary rewards, and encouraged them to *"contribute new interfaces and connectors"*(CCI5). Informants share *"though focusing primarily on the targets, some ideas could come out of the blue* [from the developer community], *and then we would evaluate them internally to understand their potential value for the platform"* (CCI5). This access to a generative community of developers facilitated immediate client support and the continuous expansion of the platform's pre-built library of interfaces (e.g., SDKs, APIs) over time. Approximately *"half of these interfaces and connectors were contributed by the community"*, enhancing the platform's generative capacity and attracting more clients and data for continued learning and improvement.

CAI-Corp's efforts to open-up its platform to the external developer community resulted in multiple feedback loops. First, the platform became more *adoptable* through diverse showcasing of its features and capabilities with a number of developers joining the developer knowledge space over the years. Second, CAI-Corp got access to the variety of natural language data through a series of multilingual CAs offered throughout the years and was able to strengthen platform's AI capability by being exposed to vast and varied data upon which platform's generative capacity would grow further. Finally, CAI-Corp got access to a generative community of developers and leveraged these external resources by extending its library of interfaces and thus functionality of the platform. Moreover, this enabled the platform to be accessed and used not only by the 'low-code' community of clients, i.e., non - technical experts that could benefit from the platform's GUI, and pre-built connectors. The platform also extended its reach to the 'pro-code' community of developers offering pre-configured code samples, and interface libraries with extensive API and SDK catalogs that could be tweaked in any possible way.

5. Discussion

Current literature has primarily viewed digital platforms as markets (Gregory et al., 2021) driven by direct and indirect network effects. However, our knowledge of specific forms of network effects, such as DNEs (Gregory et al., 2021), and how such emerge over time is limited (de Reuver et al., 2018). With the proliferation of AI platforms (Rai et al., 2019; Sundberg & Holmström, 2023), and their ability to learn and improve over time makes research on DNEs timely and relevant. Still few empirical studies have been done. Employing a generativity perspective (Thomas & Tee, 2022; Sun et al, 2023), we traced the evolution of a CAI platform over 12 years to understand and explain how DNEs emerge and expand over time in AI platforms. In addressing this research question, our study generates three key contributions.

First, we illustrate empirically how DNEs may emerge and expand over time through generative feedback loops. As observed in our study, such loops are initiated by the addition of functions that extend the generative capacity of the AI platform. While these may be added to primarily meet direct functional requirements of ecosystem actors, the new actions they afford also lead to the generation of more and increasingly diverse data that becomes available to the AI platform. Through the platform's AI capability, it is able to learn from this data and leverage it for making further improvements to the platform (Gregory et al., 2020). When these improvements lead to an increase in the platform's generative capacity, a new generative feedback loop may initiate where the scale and scope of accessible data expands, and AI capability learns and improves the platform. This demonstrates the importance of platforms' AI capability in leveraging DNEs (Gregory et al., 2020), but also how the development of AI capabilities depend on the ability of platforms to increase the scale and scope of data generated through its use over time.

The case of CAI-Corp illustrates when improvements to a platform's functionality are likely to initiate and extend DNEs and not. When CAI-Corp implemented the plug-in from the BI provider in episode 1, it indeed directly extended the platform with functions that enabled new ways for clients to visualize and analyze customer data. However, due to the way access to the data generated by use of the plug-in was restricted, no generative feedback loop could be initiated. Once it launched its own analytics function, CAI-Corp could leverage the data generated from its use to further enhance the platform, thus setting generative feedback loops in motion. This extends the notion that *"data do not need to be owned (yet accessed) to learn and improve through the use of AI"* (Gregory, Henfridsson, Kaganer et al., 2022, p. 190) by highlighting how AI platforms, by leveraging their modularity to integrate third-party applications, may restrict their own capacity to develop DNEs if integrations imply restrictions to data access.

Second, while the role of ecosystem actors as co-creators of value has been promoted as key to digital platform growth (e.g., Lehman et al., 2002), existing literature has primarily focused on the relationship between platform providers and app developers while other ecosystem actors are often overlooked. In our analysis, we observed how distinct forms of DNEs developed through two parallel processual trajectories related to both clients and external developers. The analysis shows how a platform provider may cater to both actors through tailored and gradual enhancements of a platform's generative capacity, i.e., by making it more accessible, adaptable, leveraging, and adoptable (Zittrain, 2008), and how this can initiate distinct generative feedback mechanisms that fuel DNEs. Similarly, to Eaton et al., (2015), our results indicate that extensions to the generative capacity of a platform may be the result of complex interactions between immediate and peripheral actors in the wider ecosystem. In the case presented here, clients (i.e., an immediate ecosystem actor) requested ML for their CAs at the same time as competitors from the extended ecosystem, i.e., IBM, Microsoft, and Google, introduced ML-based platforms for CA development. Both influences were key in prompting CAI-Corp to integrate ML, and ultimately both drove the development of the hybrid AI capability, thus situating the emergence of DNEs in a broader ecosystem of actors and their interactions. As their AI capability is becoming key to the core value of many digital platforms (Rai et al., 2019), digital platform research should widen the analytical scope beyond the provider-developer relationship and explore how platforms' ability to learn and improve emerges through interactions within and between several types of ecosystem actors.

While beyond the focus of this paper, we also noted how both competing digital giants and external consultancy firms may substantially influence the evolution of an AI platform. Ultimately, the AI capability of a platform depends on the ability to learn, adapt and improve, which in turn necessitates access to an increasing scale and scope of data. To that end, digital giants can often leverage their already established massive ecosystems of data generating actors when entering into the AI platform market. Therefore, smaller AI platform providers may need to balance between dependence and autonomy over time by making careful decisions about when and with whom to partner up with and at the same time ensuring that the unique value of the platform core can be maintained (Skog et al., 2021). Similarly, our data indicated that external consultancy firms had been actively driving a market demand for integrating CAI platforms with third-party systems. As a response, CAI-Corp started the development of pre-built interfaces to enhance the platform's adaptability and accessibility over time.

Third, while previous research has emphasized the role of learning and improving based on data generated by the direct use of AI platforms (Gregory et al., 2021), our study shows how DNEs may extend further when user data is complemented with additional sources. Ultimately, the scale and scope of accessible data limits and enables the development of AI capabilities, which suggests that AI platforms may benefit from attaining wide interoperability with external systems. AI platforms are often designed to operate in a wide range of business contexts, and thus expected to exchange data with diverse ecologies of both internal client systems and external web services. CAI-Corp shows how integrating with these systems can provide direct user value, but also how it may provide access to more and more diverse data that can be leveraged. To increase interoperability, CAI-Corp engaged external developers to gain both discursive inputs and actual code for the development of a multitude of interfaces. These interfaces enabled clients to connect CAs with a wider range of back-end systems, and at the same time, the platform became more attractive to developers through the availability of pre-configured code samples and interface libraries, which enabled the creation of more interfaces. These findings indicate that the generativity fueling DNEs may have characteristics distinct from, and even contradictory to, the type platform owners often strive to promote in the relationship to app developers. Rather than strictly controlling the variety of resources produced by an ecosystem through interface governance (e.g., Ghazawneh & Henfridsson, 2013), AI platforms may benefit from promoting a much wider scope and scale of resources.

In summary, the emergence and extension of DNEs in an AI platform depend on securing and extending the platform's generative capacity, leveraging generative feedback mechanisms from different actors, and accessing and relying on different types of data.

6. Conclusion

This study contributes to the understanding of how data network effects (DNEs) emerge and expand in AI platforms over time. Specifically, it presents an empirical account of initiating and expanding DNEs by following a 12-year development and evolution of a conversational AI platform. The findings highlight the importance of generative feedback loops and the platform's AI capability in leveraging DNEs. The study also emphasizes the role of ecosystem actors, including clients and external developers, in driving the development of DNEs through their interactions with the platform. Furthermore, the research expands the analytical scope beyond the provider-developer relationship, recognizing the broader ecosystem dynamics that influence the platform's generative capacity. Practical implications include the need for platform owners to actively develop AI capabilities and extend the scale and scope of data to sustain and enhance DNEs. Additionally, the study underscores the strategic considerations for the platform owners in balancing dependence and autonomy in relation to resourceful giants and the potential role of extended ecosystem actors.

This study is not exempt from limitations. Its single-case design may limit the generalizability of the findings to other AI platforms. While the case study approach is not intended to provide broad generalizations (Yin, 2018), future research could consider comparative studies across multiple AI platforms to offer more comprehensive insights into the emergence and operation of DNEs.

7. References

Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Artificial Intelligence in Organizations: Implications for

Information Systems Research. *Journal of the Association for Information Systems*, 22(2), 281–303.

Bygstad, B. (2017). Generative innovation: A comparison of lightweight and heavyweight IT. *Journal of Information Technology*, 32(2), 180–193.

Cloutier, C., & Langley, A. (2020). What Makes a Process Theoretical Contribution? *Organization Theory*, 1(1), 1–32.

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, 33(2), 124–135.

Eaton, B., Elaluf-Calderwood, S., Sørensen, C., & Yoo, Y. (2015). Distributed tuning of boundary resources: The case of Apple's iOS service system. *MIS Quarterly*, 39(1), 217–243.

Eck, A., Uebernickel, F., & Brenner, W. (2015). "The Generative Capacity of Digital Artifacts: A Mapping of the Field" (2015). *PACIS 2015 Proceedings*. 231.

Fürstenau, D., Baiyere, A., Schewina, K., Schulte-Althoff, M., & Rothe, H. (2023). Extended generativity theory on digital platforms. *Information Systems Research*, 1-25

Gawer, A. (2021). Digital platforms' boundaries: The interplay of firm scope, platform sides, and digital interfaces. *Long Range Planning*, 54(5), 1-16.

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.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2022). Data network effects: Key conditions, shared data, and the data value duality. *Academy of Management Review*, 47(1), 189-192.

Henfridsson, O., & Yoo, Y. (2014). The liminality of trajectory shifts in institutional entrepreneurship. *Organization Science*, 25(3), 932–950.

Langley, A. N. N. (1999). Strategies for Theorizing from Process Data. *The Academy of Management Review*, 24(4), 691–710.

Sundberg, L., & Holmström, J. (2023). Democratizing artificial intelligence: How no-code AI can leverage machine learning operations. *Business Horizon*s, 1-18.

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

Rai, A., Constantinides, P., & Sarker, S. (2019). Next-Generation Digital Platforms: Towar Human-AI Hybrids. *MIS Quarterly*, 43(1), iii–ix.

Skog, D. A., Sandberg, J., & Wimelius, H. (2021). How Spotify Balanced Trade-Offs in Pursuing Digital Platform Growth. *MIS Quarterly Executive*, 20(4), 259–274.

Skog, D., Wimelius, H., & Sandberg, J. (2018). Digital Service Platform Evolution: How Spotify Leveraged Boundary Resources to Become a Global Leader in Music Streaming. *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)*, 4564–4573.

Staub, N., Haki, K., Aier, S., & Winter, R. (2022). Governance Mechanisms in Digital Platform Ecosystems: Addressing the Generativity-Control Tension. *Communications of the Association for Information Systems*, 51(1), 906–939.

Sun, J., Xu, D., & Karanasios, S. (2023). How Parts Connect to Whole in Building Digital Generativity in Digital Platform Ecosystems. *ECIS 2023*, 1–11.

Thomas, L. D. & Tee, R. (2022). Generativity : A systematic review and conceptual framework. *International Journal of Management Reviews*, 24(2), 255–278.

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

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, 21(4), 675–687.

Um, S., Yoo, Y., Wattal, S., Kulathinal, R., & Zhang, B. (2013). The architecture of generativity in a digital ecosystem: A network biology perspective. *Thirty Fourth International Conference on Information Systems (ICIS)*, 1-14.

Wareham, J., Fox, P. B., & Giner, J. L. C. (2014). Technology ecosystem governance. *Organization Science*, 25(4), 1195–1215.

Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (6th ed.). SAGE Publications.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). The new organizing logic of digital innovation: An agenda for information systems research. Information Systems Research, 21(4), 724–735.

Zittrain, J. L. (2008). *The Future of the Internet and how to stop it.* Yale University Press.