Understanding KSFs in Generative Conversational AI Applications

Sarah von Bargen[1], Vittorio Molaschi[2], and Paolo Cervini[3]

Conversational AI (CAI) has been making headlines since the mid-2010s. In the last two years, as CAI and Generative AI (GenAI) have begun to combine through products such as OpenAI's GPT series and Meta's LLaMA, this hybrid has exploded in popularity.[i] Companies and individuals across industries are rapidly adopting it, finding case-specific applications for the technology. The sudden explosion of both established companies and start-ups developing products in this space leads to the question, What makes a successful generative conversational AI product?[ii]

The purpose of this case study is to provide a qualitative analysis on how companies have recently and successfully implemented generative conversational AI (GCAI). For this case study, we have chosen 4 companies to analyze: Duolingo (education), Shopify (e-commerce), Bloomberg (finance), and OpenAI (tech). After our analysis, we summarize each company's GCAI application with a Key Success Factor (KSF), which can be applied more broadly. We chose these companies for two primary reasons: not only are these digital companies with platform elements, but also they are at the top of their industries from a brand perspective, and thus provide excellent examples of digital companies which have successfully implemented GCAI. By studying companies across industries, we focus on universal KSFs, rather than industry-specific ones. We also chose companies with their primary regional loci in the United States for simplicity.

This case study is meant to give a qualitative approach to understanding some of the KSFs necessary to (1) build a CAI product, company, or platforms, or (2) integrate CAI efficiently and effectively in a company's processes. For a more in-depth, quantitative analysis, we refer you to our paper, *How does device type and product reliability affect user engagement and retention in Conversational AI? An empirical analysis using Alexa Skills data*. The purpose of developing *both* a qualitative and quantitative approach to understanding KSFs in this space is to capture effects which may be difficult to identify in one study type alone, thereby creating a thorough and encompassing approach to our analyses of this cutting-edge technology.

For managers, we provide a succinct summary at the end.

[1] Harvard Business School, Soldiers Field, Boston, MA 02163, svonbargen@hbs.edu

[2] Vocalime, Piazza Città di Lombardia 1, Milan, Italy, vittorio@vocalime.com

[3] Capgemini, Via Gattamelata 24, Milan, Italy, paolo.cervini@capgemini.com

I. Opening Summary

Although it has been around for nearly a decade, CAI (and GCAI) gained significant attention in 2022 with the advent of chatbots such as OpenAI's ChatGPT. This marked a pivotal time when chatbots, and GCAI more broadly, became part of the mainstream. While consumers had widespread access to CAI products such as Amazon's Alexa, Apple's Siri, and Google Assistant, the use cases were often limited for consumers, and difficult for companies to monetize. Thus, the popularity explosion for generative conversational AI applications such as ChatGPT led other companies, including Meta and Google, to throw their hats in the ring with similar products such as LLaMA and Bard/Gemini.

The appeal of GCAI to consumers

As CAI applications transition to GCAI, potential B2C applications exploded. One of the core questions was: Would consumers actually use it? If yes, would they prefer audio- or visually-focused engagement? In a recent study, researchers at Stanford University and the University of Notre Dame performed a series of experiments on two online platforms to determine whether users valued spoken or written recommendations more. Not only did the researchers discover that the audio recommendations were followed more than the written/read ones, they also identified that users responded equally to human and computer-generated voices. The researchers believe that because users can re-access the written recommendations more easily than spoken ones, users place more emphasis on the audio recommendations. "The ephemerality of auditory messages evokes a sense of relative urgency…lead[ing] individuals to exhibit a speeded-up reliance on early cues."[iii,iv] However, the type of GCAI product has a direct effect on how users prefer to engage with it . In our separate research paper, we identify that users prefer visually-focused devices in the long run for game-style Alexa skills which contain both an audio-focused and visually-focused engagement option.[v] Thus, it is essential for managers, entrepreneurs, and developers to understand their product and audience, and how users would want to engage with their GCAI product.

Business Models

In tandem with GCAI's nature, companies are developing three main categories of business models: horizontal, vertical, and hybrid models. Horizontal models are how we traditionally think about chatbots; they can be applied across company types, industries, etc. GPT-4 and Meta's LLaMA are classic examples of this model. Vertical models have the opposite features. They cater to a specific industry, and sometimes a particular company. For example, BloombergGPT was trained on public and in-house financial data, for their Bloomberg Terminal users to have a more in-depth and engaging experience. BloombergGPT acts similarly to a personal assistant for customers using the terminal. Finally, hybrid models combine the previous two models through mechanisms such as plug-ins. Although the plug-in can be acquired or developed by a wide range of industries, they also can be *customized* to a specific industry's or

company's needs. OpenAI's recent announcement surrounding their new store with these plug-ins is one of the forerunners in this hybrid model.

In many cases, horizontal and hybrid models are more likely to be consumer-facing than vertical models. From a monetization standpoint, horizontal and hybrid models show great promise for driving revenue increases. On the other hand, vertical models provide cost-cutting opportunities, by streamlining back-office processes, customer support, data analysis and aggregation, etc. In terms of monetization, vertical CAI models can potentially be utilized for more efficient product/service development.

I. Duolingo: Using GCAI to build ground-breaking products

Our first company is Duolingo, a language learning SaaS transaction platform[vi] which was founded in 2011 by Luis von Ahn and Severin Hacker, a professor and a graduate student at Carnegie Mellon University. von Ahn had a successful history of designing computer-based programs, including reCAPCHA, used to identify whether a user is a human or robot on a website. Between 2009-2016, they created "new computational algorithms to better understand how connections are formed and how information self-organizes on social networks," which would later become essential to introducing CAI into their products. By 2016, the company had raised $83 million in VC investments.[vii]

In July 2021, Duolingo debuted its IPO at a valuation between $3 billion to $6.5 billion, with share prices closing higher than expected on the first day of trading.[viii] By the end of 2023, both Duolingo's revenue and stock prices were exploding in growth. In February 2024, Duolingo's market cap was around $7.52 billion.

AI and CAI in Duolingo's Products

Since 2018, Duolingo had actively implemented artificial intelligence in their products, but the platform accelerated its push into GCAI in March 2023, with the introduction of Duolingo Max. This new subscription service offered users more expanded features to assist them in their language-learning journey, while also providing Duolingo a way to run A/B tests on their new CAI products. Duolingo Max included two new features, which were built in partnership with OpenAI: "Explain my Answer", and "Roleplay." Built on GPT-4's technology, these features provided users with a more-enhanced learning experience.

Specifically, "Explain my Answer" provided feedback to users, explaining why their answer was wrong and how to correctly think about the grammar, vocabulary, etc. Users were able to text with the chatbot to ask deeper, more personalized questions to understand not only how, but why they made a mistake. The other new feature, "Roleplay," gave users a way to practice conversational language by talking with a chatbot, which had been trained to converse as a human would, visualized as one of the iconic characters in the app. These conversations

were themed, helping users practice specific vocabulary within scenarios such as ordering at a restaurant or making vacation plans.

Figure 1. Example of "Explain my Answer" and "Roleplay" in use

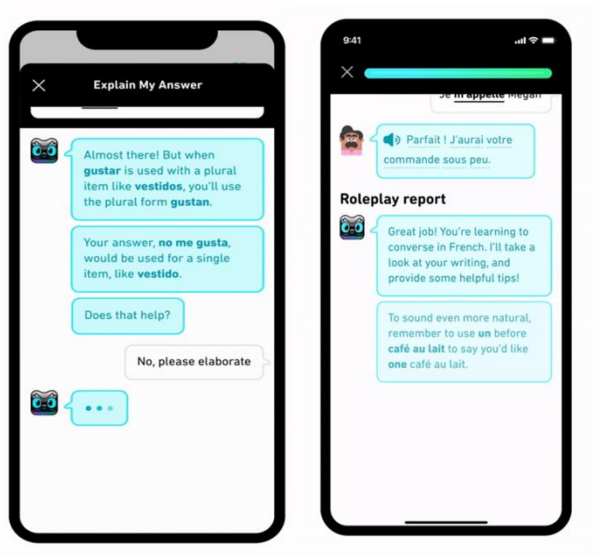


Photo source: Duolingo

Duolingo Max was initially debuted in two languages (English speakers learning French and Spanish) in the U.S., Great Britain, Ireland, Australia, Canada, and New Zealand. In the early to mid 2023, Duolingo utilized human moderators and translators to ensure the content was accurate, but by the end of 2023, around 10% of these contract-based moderators were terminated.[ix]

Until 2023, Duolingo had never been profitable. However, by Q2 2023, the company generated net income for the first time in the company's history, and maintained profitability through the third quarter. By Q3 2023, Duolingo broke another record with a DAU/MAU ratio of 29%.[x] Analyst outlooks also improved for the platform, with many citing the introduction of Duolingo's AI products as one of the primary drivers for their positive outlook on the company's profitability.

Key Takeaway: Focus on Design to Enhance User Engagement with CAI

One of the most notable strategic decisions by Duolingo was emphasizing their design, and repeatedly testing variants of the design to ensure their product would be successful. Not only did this solidify their brand identity, but also made their products fun, consistent, and engaging for users. It also *seamlessly* integrated CAI into their products. Duolingo's products were

consistently easy to use, as frequent glitches or latency were not reported (it even sports a 5 star review from PCMag in 2023).

II. Shopify: Helping two sides of a market with GCAI

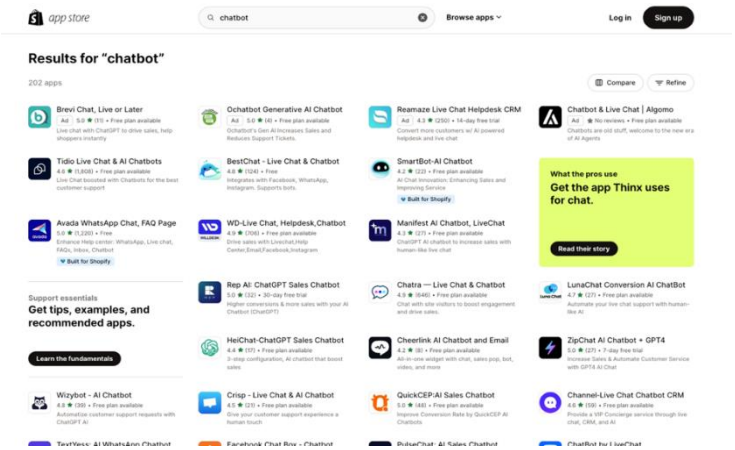
Of the many companies implementing CAI and GCAI into their business strategies, Shopify has been one of the most thorough and seamless. In this section, we will focus on two of their GCAI applications: Shopify's App Store which, among other things, offers entrepreneurs a way to use a third-party app to build a personalized chatbot for their online store, and Shop.app, a transaction platform built by Shopify which connects stores made on Shopify's software with customers. These buyers can use a Shop.app chatbot to assist them with their purchase.

1) Shopify's App Store

Shopify is a transaction platform, meaning they specifically focus on facilitating transactions between 2 or more sides of the market. The App Store is a classic example of this dynamic in action, where app developers are matched with entrepreneurs building stores in Shopify. According to the company, 73% of customers would rather purchase from a store which provides tailored shopping experiences.[xi] A well-designed and implemented chatbot provides online stores a low-cost way to personalize their buyer's experience.

This trend of chatbots as part of the shopping experience appears to be gaining traction globally. PricewaterhouseCoopers released a report on consumer attitudes towards "chatbots as online shopping assistants in 2023." In this study, PwC surveyed 8,975 consumers in 25 countries and territories about their interest in customer-facing chatbots and online shopping. 44% of consumers said they would utilize a chatbot for "searching for product information before I purchase (e.g. product details, recommendations, comparisons)", and 31% were interested in a chatbot which "customize[d] offerings."[xii]

Figure 3. Shopify's App Store, where entrepreneurs can pick their chatbot provider

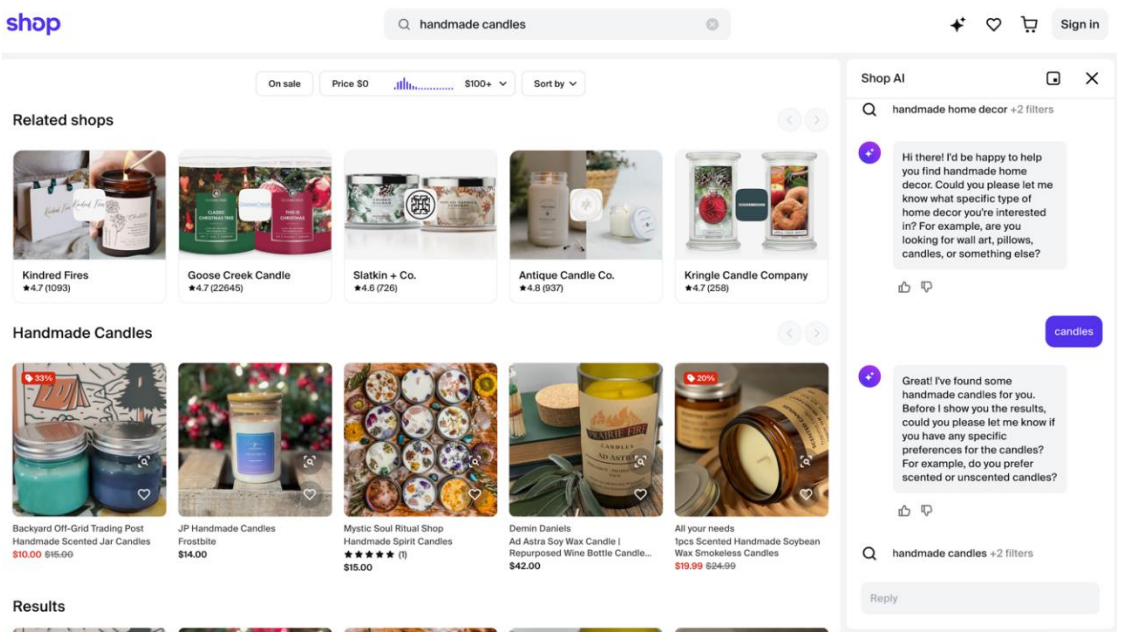


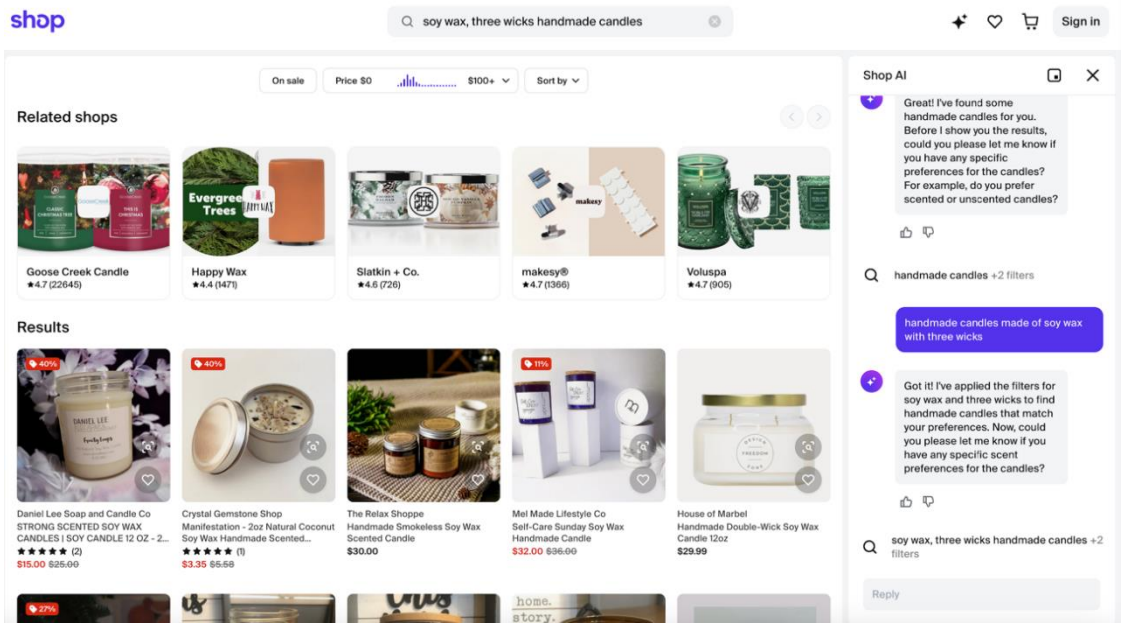
Source: Shopify

2) Shop.app

In 2023, Shopify released a new version of their online marketplace, called Shop.app, which incorporated a chat-based shopping assistant.[xiii] Customers could enter the products they were looking for into the search engine, and a chatbot would appear which would customize their shopping experience. As the conversation expands, the options available to the user would change, becoming more tailored to their needs.

Below are examples of a conversation between a shopper and the chatbot:





Source: Shopify

By February 2024, Shopify released a similar chatbot, called Semantic Search, for entrepreneurs to integrate into their own shops.[xiv, xv] Utilizing predictive search, the goal for this chatbot was to tailor a shopper's needs while also improving the user's search experience. This chatbot was similar to the chatbot described above, but could be installed into a specific entrepreneur's shop so it was more tailored to each store owner. Semantic Search was designed to understand a customer's intuitive thinking when shopping for items, even if a seller did not tag items with specific keywords. For example, if someone typed "Valentine's Day dress," images of heart-themed pink and red dresses would appear, regardless if the vendor had utilized the words "Valentine's Day" in the description.

Shopify's focus on improving customers' search engine experiences was rooted in turnover rates, as customers demonstrated a lack of patience for technology which did not work properly or meet their needs. One study discovered that 69% of customers immediately use the search engine bar when visiting an online store, and 80% had left an online retailer in the past due to a "poor search experience."[xvi] Shopify's strategy to move step-by-step with their various product releases exemplifies a delicate balance faced by many digital platforms: release a product too fast, and the GCAI may have too many issues, leading to lost customers; too slow, and the platform could lose their first-mover advantage. By taking this step-by-step approach, Shopify has helped mitigate some of the risks of this timing balance, while gaining customer feedback for the next products' launches.

<u>Key Takeaway</u>: Tailor to your users' needs and ensure your generative conversational AI product works efficiently from the start. People will not wait too long for yours to work; they will move on to another system.

III. Bloomberg (via BloombergGPT):

By 2023, most GCAI products were for generic applications. However, some companies began making niche generative conversational AI products, specific to their industry's or company's needs. To cater to the financial industry's needs, Bloomberg built and trained a 50 billion parameter LLM known as BloombergGPT. According to David Rosenberg, head of the machine learning strategy team in the Office of the CTO, Bloomberg split the training data between "general training data" and "Bloomberg-specific financial training data" called FinPile. FinPile is comprised of "news data, financial filings (company filings, press releases, transcripts of calls), and a broad collection of financial data including tables and charts," and contains data from 2007 to the present. Bloomberg did not perform additional specific processing for the FinPile dataset, but "when it was already extracted, we used it to train [the model]".[xvii]

One of the first applications of BloombergGPT was internal. As Bloomberg explained in their press release:

> "Recent advances in Artificial Intelligence (AI) based on LLMs have already demonstrated exciting new applications for many domains. However, the complexity and

> unique terminology of the financial domain warrant a domain-specific model. BloombergGPT represents the first step in the development and application of this new technology for the financial industry. This model will assist Bloomberg in improving existing financial NLP tasks, such as sentiment analysis, named entity recognition, news classification, and question answering, among others. Furthermore, BloombergGPT will unlock new opportunities for marshalling the vast quantities of data available on the Bloomberg Terminal to better help the firm's customers, while bringing the full potential of AI to the financial domain."[xviii]

In addition, because Bloomberg integrated their GPT into their terminals, users had access to the product, and could perform tasks such as sentiment analyses directly using BloombergGPT.[xix]

<u>Key Takeaway</u>: Differentiate your GCAI product. In Bloomberg's case, this meant offering one of the first finance-specific GPTs to their Terminal users. Whether differentiation means being industry-specific, region-specific, etc., know what your customer is looking for in a GCAI use case, and make that one of your product's unique strengths.

IV. OpenAI's GPT store

OpenAI is one of the most iconic and popular companies in the AI space. While we could write an entire case study on this company alone, for our purposes we are focusing on one specific segment of the company; their recent GPT store.

Founded in 2015, OpenAI exploded in growth in late 2022 with the immense popularity of ChatGPT. In January 2024, OpenAI launched a new hybrid service which blended the benefits of their horizontal and vertical products. The GPT store, where users could access custom chatbots, was available to ChatGPT Plus subscribers, with the monthly fee set at $20. Some of the GPTs available included "AllTrails personal trail recommendations, a Khan Academy programming tutor, a Canva design tool, and a book recommender." Users could also develop their own GPTs, regardless of their programming skill level, and offer the chatbots in the store. However, at the time of the launch, OpenAI did not have a plan for how companies would generate revenue from the store. As Sam Altman, CEO of OpenAI explained, "Revenue sharing is important…We're going to pay people who build the most useful GPTs a portion of our revenue." Nonetheless, the GPT store had 3 million chatbots available at the time of its launch.[xx]

However, by March 2024, the GPT store was already facing challenges and complaints from users. One common problem was monetization; the store was only available to ChatGPT's paying subscribers, which limited the number of people who could purchase the bots. Other users noted problems with spammy products and copyright violations, with some developers

creating products which can mimic famous celebrities and/or companies without their permission. Making matters more complicated is that OpenAI may not be liable for copyright infringement due to the Digital Millennium Copyright Act, "which protects it and other platforms (e.g. YouTube, Facebook) that host infringing content so long as those platforms meet the statutory requirements and take down specific examples of infringement when requested."[xxi]

These series of problems in the GPT Store had caused some developers to question whether OpenAI was ready for a store, given the number of glitches out of the starting gate. Additionally, creators are frustrated with the lack of support and moderation from OpenAI. As one developer explained, "People feel like [OpenAI] has abandoned the GPT store."[xxii]

<u>Key Takeaway</u>: Having a popular platform requires maintenance, even with an exceptionally successful GCAI product such as ChatGPT. Solid maintenance nurtures trust with developers and customers, one of the most essential elements of a successful platform and platform ecosystem.

V. Key Managerial Takeaways

While GCAI can assist in user engagement, monetization, and efficient internal operations, how to achieve these goals is still unclear for some businesses. Various companies are hesitant to adopt GenAI, CAI, and GCAI, primarily due to unfamiliarity with using the technology, or other concerns such as privacy. Even OpenAI is dealing with hesitation from potential users. Executives at companies which were deciding whether to build GPTs for OpenAI's new store expressed concerns over publishing on the store, including "[their own] limited resources to explore the store's capabilities, as well [as] its lack of defined audience and revenue growth opportunities." (This last reason likely alludes to OpenAI's unclear model for monetization for businesses at the time of the survey.) As one executive asked, "Ultimately, how much new audience [or] money is this going to generate?"[xxiii]

However, many early adopters of GCAI are already seeing the benefits; building higher user engagement, monetizing new market segments, optimizing customer service, and distinguishing themselves as a leader in their field due to more seamless and tailored experiences for their customers. GCAI also offers new opportunities for companies to make a first move in gaining customers, something especially vital for platform companies, which often depend on first moves to develop user bases at rapid speeds. This first-mover advantage was the motivation for Ingenio, "an online marketplace for knowledge exchange,"[xxiv] which developed three GPTs for the OpenAI store within the first month. As Josh Jaffe, Ingenio's President of Media explained, "Sometimes if you wait too long, you really face an uphill battle in gaining visibility in these platforms [referring to the GPT store]."[xxv]

As GCAI develops, potential applications are expanding rapidly across areas including marketing and sales, manufacturing, and product development. With this expansion of applications, the KSFs for successfully applying GCAI will also morph and adjust as companies find groundbreaking ways to integrate the technology into their regular operations. For now, we have detailed several KSFs throughout this case study which apply to current GCAI applications, and detail them again below.

Develop an easy to use product, which is consistent and correct.

Invest in developers, especially if you are a platform, and ensure your company is producing high-quality, reliable products for users.

Match the GCAI product with the hardware; do users prefer a visually-focused or audio-focused engagement with your GCAI product?

Ensure your product works well from the beginning; users often have little patience for technology, and having a faulty product in the start can create a perfect opportunity for your competitors. At the same time, waiting too long to release a product can result in a lost first mover advantage. Therefore, it is essential to balance the timing.

Maintain your platform.

Maintaining your platform fosters trust with customers and developers. It also acts as a preliminary defense against potential competitors who can come in later and “build it better”.

Differentiate your CGAI product.

Whether differentiation means being industry-specific, region-specific, etc., know what your customer is looking for in a GCAI use case, and make that one of your product's unique strengths.

Endnotes

[i] According to IBM, conversational AI "refers to technologies, such as chatbots or virtual assistants, that users can talk to." Generative AI "refers to deep-learning models that can take raw data…and 'learn' to generate statistically probable outputs when prompted." (see: https://www.ibm.com/topics/conversational-ai and https://www.ibm.com/topics/artificial-/intelligence#:~:text=Generative%20AI%20refers%20to%20deep,statistically%20probable%20outputs%20when%20prompted.)

[ii] For the purposes of this paper, the term "product" can refer to either a product or service.

[iii] https://journals.sagepub.com/doi/10.1177/09567976221106349

[iv] *Which Are More Effective: Spoken or Written Recommendations?* HBR, September-October 2023.

[v] Sarah von Bargen, Vittorio Molaschi, and Paolo Cervini, "How does device type and product reliability affect user engagement and retention in Conversational AI? An empirical analysis using Alexa Skills data."

[vi] We use the term "transaction platform" as defined by Cusumano, Gawer, and Yoffie in their 2019 book "The Business of Platforms". Generally speaking, transaction platforms generate revenue by connecting one or more sides of the market. In the case of Duolingo, they connect users with each other via their ranking system, which allows people to compare their learning progress with other learners on the app. With the advent of GenAI, there is also discussion now on whether machine-driven network effects will be as effective as human-driven ones.

[vii] Masson-Forsythe, Heather. "NSF gave Duolingo its Wings!" *U.S. National Science Foundation*, Dec 22, 2023. https://new.nsf.gov/science-matters/nsf-gave-duolingo-its-wings#:~:text=Duolingo%20was%20founded%20in%202011,by%20NSF%20at%20the%20time.

[viii] Wang, Echo and Nishant, Niket. "Duolingo enters 'major leagues' with $6.5 bn valuation in strong debut." *Reuters*, Jul 28, 2021.

[ix] De Vynck, Gerrit. "Duolingo cuts workers as it relies more on AI." *The Washington Post*, Jan 10, 2024, updated Jan 11, 2024. https://www.washingtonpost.com/technology/2024/01/10/duolingo-ai-layoffs/

[x] Barclays, "Duolingo, Inc.: Solid Across the Board, Math & Music about to crank things up."

[xi] Lizzie Davey, "What is a Chatbot? Types of Chatbots and How They Work." *Shopify*, Nov. 2, 2023, https://www.shopify.com/blog/chatbots

[xii] Daniela Coppola, "Share of consumers interested in chatbots as online shopping assistants in 2023." PriceWaterhouse Coopers and Statista, June 2023.

[xiii] Yiwen Lu, "My Not-So- Perfect Holiday Shopping Excursion with A.I. Chatbots." *The New York Times*, Dec 14, 2023, accessed March 14, 2024.

[xiv] This feature was "available to Shopify Plus customers with fewer than 200,000 products and English-language storefronts." Source: Inc.com, https://www.inc.com/rebecca-deczynski/shopifys-new-ai-search-feature-helps-sellers-recommend-right-products-to-customers.html

[xv] "Modifying search." *Shopify Help Center*, https://help.shopify.com/en/manual/online-store/search-and-discovery/search#semantic-search, accessed March 14, 2024.

[xvi] Rebecca Deczynski, "Shopify's New AI Search Feature Helps Sellers Recommend the Right Product to Customers." *Inc.com*, Feb 1, 2024, https://www.inc.com/rebecca-deczynski/shopifys-new-ai-search-feature-helps-sellers-recommend-right-products-to-customers.html, accessed March 14, 2024.

[xvii] "BloombergGPT – an LLM for Finance with David Rosenberg – 639." *The TWIML AI Podcast with Sam Charrington*, July 24, 2023. https://www.youtube.com/watch?v=byCe7-c84d4

[xviii] https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/

[xix] https://ambcrypto.com/blog/how-to-use-bloomberg-gpt/

[xx] Stephen Shankland, "OpenAI's GPT Store Now Offers a Selection of 3 Million Custom AI Bots." Jan 10, 2024, https://www.cnet.com/tech/computing/openais-gpt-store-now-offers-a-selection-of-3-million-custom-ai-bots/, accessed March 15, 2024.

[xxi] Kyle Wiggers, "OpenAI's chatbot store is filling up with spam." *TechCrunch*, March 20, 2024, https://techcrunch.com/2024/03/20/openais-chatbot-store-is-filling-up-with-spam/, accessed March 22, 2024.

[xxii] Stephanie Palazzolo, "OpenAI's Chatbot App Store Is Off to a Slow Start." *The Information*, March 18, 2024, accessed April 1, 2024.

[xxiii] Sara Guaglione, "Why publishers are hesitant to add their chatbots to OpenAI's GPT Store." *Digiday*, Jan 26, 2024, https://digiday.com/media/why-publishers-are-hesitant-to-add-their-chatbots-to-openais-gpt-store/, accessed March 16, 2024.

[xxiv] "Ingenio LLC." *Bloomberg, https://www.bloomberg.com/profile/company/1211273D:US,* accessed March 16, 2024.
[xxv] Sara Guaglione, "Why publishers are hesitant to add their chatbots to OpenAI's GPT Store." *Digiday,* Jan 26, 2024, https://digiday.com/media/why-publishers-are-hesitant-to-add-their-chatbots-to-openais-gpt-store/, accessed March 16, 2024.