Personalization in Language learning platforms: How does Duolingo combine AI, gamification, and personalization strategies to shape user experiences while navigating privacy concerns?

Bachelor's Thesis, Copenhagen Business School, 2025
Bachelor of Science in Business Administration and Digital Management

Author: Elitsa Stoimenova (Student No. 162042)
Supervisor: Margrethe Lyngs Mortensen
Date: 13th of May 2025
Pages: 37
Characters: 84,500

Abstract

This study focuses particularly on how the emergence of digital platforms has reshaped the educational industry. It does so by conducting an in-depth analysis of users' perspectives on Duolingo's integration of AI-driven personalization, gamification mechanics, and privacy considerations.

From a digital platform standpoint, user interactions are at the heart of value creation. While the root of every user-platform interaction lies in the interface, it is essential to pinpoint where its design falls short. Nielsen's Human-Computer Interaction heuristics (1994) support this, providing an overview of system visibility, minimizing cognitive load, and error prevention. However, when addressing interface frictions, it is essential to first grasp the underlying algorithms that power adaptation. Thus, by employing a combined framework of Human-Computer Interactions and Natural Language Processing algorithms, this study aims to diagnose users' interactions with Duolingo, ultimately informing design refinements that make adaptability apparent and pedagogically coherent.

As the primary focus of this study is a case company, the strategy for the case study method is adopted. The methodological approach is primarily pragmatic, utilizing a mixed-method approach to investigate the findings from a specifically curated online survey. The data collection is done over a week, while a convenience sampling method targeted English-speaking university students, who were likely to have previous experience using the app. The mixed-method analysis revealed four key friction points: a gap between users' perceived and actual personalization, the dominance of habit-forming streaks over genuine skill development, the user's dissatisfaction with the app's one-size-fits-all adaptation approach, and the mistrust around unjustified data collection.

Finally, during the discussion, these tensions were mapped to actionable design improvements. Ultimately, it is concluded that explaining and visualizing personalization should be Duolingo's main priority.

Table of Contents

Introduction

Over the last few decades, we have become witnesses to the rise of the Internet of Things (IoT) and digital platforms, which have triggered what Schumpeter once described as the process of "creative destruction" (Schumpeter, 1942). He defines this process as the reconfiguration of markets under the force of innovation, where old models are disrupted by new and more efficient structures, allowing them to emerge as market leaders. The rise of digital platforms has fundamentally altered traditional industries, shifting the focus from linear value chains towards dynamic systems where users, data, and algorithms interact simultaneously. Their emergence has signaled not only a technological but also a structural transformation in how economies operate. Digital platforms have taken a dominant position in various sectors, including finance, retail, entertainment, and education, through their ability to act as facilitators of interactions. As Parker et al. (2016) argue in their book, "Platform Revolution", this constructs a new perspective on how value is created, shifting the center of it from product delivery to user interactions. The key innovation behind platform dominance lies in its reformed architecture. Parker et al. (2016) state that digital platforms enable the "pull, facilitate, and match" cycle (p.48). They attract users (pull), foster interactions (facilitate), and create personalized experiences through algorithms (match). Thus, value is no longer produced linearly through standardized products that users must adapt to. Instead, value becomes dynamically personalized to fit each user's preferences and behaviors.

However, as described in "Platform Revolution", platforms "do not control value creation; instead, they create an infrastructure in which value can be created and exchanged, and lay down principles that govern these interactions" (Parker et al., 2016, p.50). These principles guide who can participate, the types of interactions enabled, how users are matched with other users or content, and how data flows through the system. This turns platforms from value producers into orchestrators of conditions in which users, algorithms, and interfaces interact to generate outcomes. This new platform architecture is built upon continuous data collection and feedback loops. Platforms rely on machine learning, and real-time data analytics to adapt content, timing, and interface behaviors. In this new model, the user is not only the recipient of personalized experiences but also the source of behavioral data, which informs future interactions. As the user becomes an undeniable part of the new value creation model,

questions about algorithmic transparency and ethical use are bound to arise. Thus, presenting new challenges that are not encountered in traditional models of value creation.

Many industries have undergone significant changes following the disruption of digital platforms, but one industry that is currently experiencing this disruption is education. Digital platforms have reshaped the way knowledge is delivered and accessed, completely revamping how education can be structured. The process of acquiring knowledge is no longer confined to a classroom, through syllabi, fixed timetables, and hundreds of pages of material to read. Knowledge can now be developed on the go from any location in the world, personalized to the user's pace, likes, and learning style. In the field of educational platforms, language learning platforms stand out as a particularly illustrative example of this transformation. Language learning is suitable for mobile formats due to its modularity, repetition-based structure, and the need for continuous practice (Sharples et al., 2005). These platforms aim to create a space where users are incentivized and motivated to acquire new knowledge by transforming the typical learning process into a gamified, interactive environment. Among the many Mobile-Assisted Language Learning platforms (MALL), Duolingo stands out as the most prominent and widely adopted platform.

As a platform that combines personalization, gamification, and algorithmic design (Duolingo, n.d.), Duolingo exemplifies the evolution of digital learning tools in the context of platform-based value creation. With over 500 million users (Duolingo, n.d.), the platform is built on a model that shapes engagement patterns and personalizes the user experience through Natural Language Processing (NLP) and gamification (Henry, 2025). As a facilitator rather than a controller of value, Duolingo presents a clear case through which the platform logic can be explored in an educational setting. Their personalization methods bring forward the concerns about using AI for personalization and algorithm transparency. This study aims to shed light on how users perceive Duolingo's personalization methods, focusing on their effectiveness and the trade-offs users are willing to accept between autonomy and personalization. Formulating the following research question:

'How does Duolingo combine AI, gamification, and personalization strategies to shape user experiences while navigating privacy concerns?'

Case description

This section of the paper presents Duolingo, providing a brief overview of its structure, key features, and relevance to this study's research. It aims to highlight the platform's core methods and strategies, which subsequently forms the foundation of the theoretical framework deployed in this paper.

Duolingo was founded in 2011 by Luis von Ahn and Severin Hacker and has since become one of the world's most popular and widely downloaded language learning apps (Duolingo, n.d.-c). With over 500 million users, Duolingo has quickly become the most recognizable market leader in the online education sector. The app is built on the freemium model, where users can enjoy free access to language learning courses with ads, or opt for a paid subscription option (Duolingo Super) with no ads (Duolingo, n.d.-b). Duolingo employs five core practices in its approach toward building the user experience: learning by doing, continuous improvement, personalization, evidence-based practices, and gamification (Duolingo, n.d.). The platform offers an interactive, task-based experience that enables users to learn by directly applying their language skills in context, sometimes in unusual situations, because they are “memorable and more fun to learn” (Duolingo, n.d.). Tailoring the user experience to individual progress levels facilitates continuous improvement and personalized learning. Which they are able to provide due to the collection and storing of data from their millions of users (*Duolingo*, n.-d.). Additionally, gamified features such as XP points, streaks, and leaderboards boost motivation and encourage habitual engagement.

Duolingo showcases core principles aligned with user-centered design through its dedication to evidence-based practices and personalization. In line with this approach, the platform adheres to Human-Computer Interaction (HCI) principles, emphasizing reduced cognitive load and rapid task completion (Nielsen, 1994). Duolingo’s design choices, with a clean, colorful, and minimalist interface, support these claims. It aims to focus user attention on a singular interaction at a time, improving attention retention and learning flow. The lessons are short and repetitive, purposefully minimizing the load experiences in a single session, sustaining user engagement. It uses visual checkmarks, sound cues, and animations (Duolingo, n.d.) to reinforce correct answers, creating a visual feedback structure that users become hooked on. Duolingo is following a gamified cycle of “challenge - success - prize”

described by Zichermann and Cunningham (2011), which has been proven to promote dopamine production in the brain, thereby increasing users' desire to play, or in this case, learn.

To provide a personalized and adaptive experience, Duolingo leverages the power of Natural Language Processing (NLP) and, more recently, Large Language Models (LLMs) (Henry, 2025). In its nature, LLM is a natural language processing model that allows Duolingo to interpret both written and spoken data. With its adoption, the app can now accept grammatical variations and semantically similar answers, further encouraging users to focus on meaning rather than guessing exact word formations. This reflects a shift from exact answer matching to a more context-aware evaluation, where users are still rewarded for partially correct answers and are continuously encouraged to engage in a positive learning experience. The real-time corrections and grammar explanations serve as a means to close the feedback loop described in HCI principles (Nielsen, 1994), where immediate system feedback and user control are key actors in effective interaction.

Considered as a whole, Duolingo presents a compelling case for how educational platforms are reshaping the well-established concept of learning. Leveraging adaptive user-centered design and data-driven interactions, Duolingo offers a gamified experience that helps users learn in a positive reinforcement environment (Duolingo, n.d.). However, the platform's use of analytical algorithms, AI, and data collection for future improvement raises concerns about behavioral influence. Duolingo's ability to constantly adjust and improve based on the collected user data illustrates the power of algorithmic systems to shape user behavior, serving better retention and engagement goals. This presents the hypothesis that Duolingo should be seen not merely as an educational platform, but also as a behavioral one. The platform becomes a space where interactions are tracked and subtly directed, opening a new perspective on how users perceive their privacy. These dynamics make Duolingo a suitable case for exploring the tension between personalization, privacy, and user autonomy.

Literature review

The following literature review section is curated based on the existing research surrounding the topic of digital language learning. It introduces concepts such as

Mobile-Assisted Language Learning (MALL) (Sharples et al., 2005), the underlying principles of gamification in education, and user motivation within digital learning. This section examines case studies, conceptual frameworks, and gamification models, intending to contextualize Duolingo's approaches within the back-end of platform–based education.

The concept of Mobile-Assisted Language Learning (MALL) emerged in the early 2000s with the research of Sharples, Taylor, and Vavoula (2005). During the initial research into the topic, multiple definitions were assigned to MALL, depending on whether the primary mobility falls onto the user or the device itself (Burston, J., 2014). According to Sharples et al. (2005), the concept should be examined from the perspective of the learner being mobile, rather than the technology. On the other hand, Kukulska-Hulme & Shield (2008) introduce the idea of 'mobile learning' as: "learning mediated via handheld devices and potentially available anytime, anywhere" (Kukulska-Hulme & Shield, 2008, p. 273), thus focusing on the technology, not the learner. To put an end to this dilemma and define the concept of mobile learning in the way it is perceived today, Palalas introduced his opinion: "MALL can be defined as language learning enabled by the mobility of the learner and... portability of handheld devices..."(2011, p.76-77). This is also the approach Burston (2014) adopted in his research on the pedagogical challenges of MALL.

Burston (2014) defines the four central pedagogical tensions: flexibility vs. consistency, micro-learning vs. depth of processing, technological adaptivity vs. pedagogical coherence, and engagement mechanics vs. education values. In his research, he synthesizes findings from multiple studies on the tension between flexibility and engagement, illustrating the struggle users experience to stay consistent without structured guidance or social pressure. He then further criticizes the micro-learning activities exhibited by MALL (e.g., flashcards or brief quizzes), pointing out that such activities predominantly support surface-level learning. According to him, adaptive algorithms without a pedagogical framework risk fragmenting the learning trajectory, making it difficult for learners and instructors to track progress against curricular goals (Burton, 2014). Finally, he touches upon the gamification features (e.g., badges, points, and leaderboards) and their importance to the learner. Although he acknowledges their value, he also highlights that over-emphasizing such features can distract learners from their language learning goals, turning the process into a game.

As already established, gamification plays a critical role in creating successful digital learning platforms. Gamification itself is not concerned with the process of designing a game; instead, its primary focus is to provide engagement, commitment, and motivation between focused situations and individuals (Topal & Karaca, 2021). In the context of education, gamification aims to dissect the elements of computer games that users find engaging and adopt them in an educational environment (Topal & Karaca, 2021). The book 'Research Anthology on Developments in Gamification and Game-Based Learning' (2022) points to the most widely applicable game features in education: fun, narrative, rules and rewards, challenges, motivation, engagement, and gameplay. Furthermore, Werbach and Hunter's 'Gamification Toolbox' (2012) defines how these features can be interpreted into dynamics, mechanics, and components. According to them, dynamics represent the narrative, emotions, constraints, and relationships. The mechanics are the challenges, competitions, feedback, resource acquisition, rewards, and win statuses, essentially the elements that motivate users. While the components represent the design objects users interact with: achievements, badges, and avatars.

From this review, the study identifies two significant gaps in the current literature. Firstly, while Burston (2014) acknowledges micro-learning adaptive environments as a double-edged sword, his research predates the incorporation of advanced AI. Consequently, there is a need for an updated perspective on how AI-driven personalization (particularly Natural Language Processing and Large Language Models) influences the pedagogical challenges he outlines. This study aims to fill that gap by assessing whether NLP/LLMs enhance coherence or further fragment the learner's trajectory. Secondly, while existing gamified frameworks, e.g., Werbach and Hunter's 'Gamification Toolbox' (2012), map out the design elements that can motivate users, there remains a lack of research connecting theory to real-life user experiences. Furthermore, this study aims to shed light on the little-discussed subject of collecting personal information for the design of AI-personalization and gamified feedback loops.

Theoretical framework

This study draws from a theoretical framework grounded in Human-Computer Interactions (HCI) and Natural Language Processing principles. These theories have been

specifically chosen because they collectively address both Duolingo's user-facing design features and the algorithmic processes behind personalization. This combined framework provides the analytical tool to evaluate the interplay between behavioral design, personalization, and trust in data-driven systems.

Human-Computer Interactions

Human-computer interactions emerged as a research area in the early 1980s (MacKenzie, 2012). It was established as a computer science area that embraced cognitive science and human factors engineering. However, with the rapid development of technologies, the field redirected its focus from analysing specialized cases to examining everyday users and their interactions with complex digital systems. In its core, HCI focuses on the interfaces between users and systems, exploring the design and use of computer technology (MacKenzie, 2012). In the context of this study, HCI provides a fundamental theory to analyze how Duolingo's interface, motivational systems, and personalization features impact user engagement and perception. Central to the modern understanding of HCI principles are the usability heuristics developed by Jakob Nielsen in 1994. They provide a framework for evaluating digital interfaces, emphasizing essential features such as: visibility status, user control, consistency, error prevention, and minimal cognitive load.

Firstly, Nielsen (1994) emphasizes the importance of informing the users. According to him, the system should always notify the users about what is happening through appropriate and timely feedback. This ensures that users learn from the outcome of their prior actions and can adjust their expectations toward the next steps. The second heuristic principle acknowledges the importance of creating a match between the system and the real world. It discusses the design and how it should reflect the user's language and use concepts, words, and phrases similar to real-world situations. Thirdly, Nielsen (1994) highlights the importance of user control and freedom. Users should be allowed to undo and redo actions, creating a feeling of navigation control and freedom of exit. Consistency and standards are marked as a fourth heuristic. Users should be provided with a uniform language, layout, actions, and feedback across the platform. Users' minds should not doubt whether the exact words or actions would result in different interactions. This is outlined as a measure to reduce the cognitive load and increase efficiency. As a fifth heuristic, Nielsen (1994) puts error prevention. According to

him, systems shouldn't rely on reactive error messages, but rather be designed to minimize the likelihood of errors occurring proactively.

Furthermore, Nielsen (1994) emphasizes the importance of recognition rather than recall. This essentially represents the idea that users should be provided with the opportunity to recognise content instead of being forced to remember said information. This heuristic is grounded in psychological research, which points to recognition being more reliable than recall as it reduces the cognitive load. Flexibility and efficiency of use are next on the list of essential principles according to Nielsen (1994). This heuristic reflects how important it is for interfaces to adapt their interactions based on the user. It recognizes that users in different stages should be guided differently - more structured and educated steps for beginners, while providing more advanced challenges for long-term users. Another aspect central to reducing cognitive load is the minimal design Nielsen (1994) advocates for. According to him, unnecessary complexities should be avoided, and primary tasks and actions should be represented clearly. The ninth heuristic addresses the importance of timely recognition, diagnosis, and recovery from errors. It is tightly connected to the first principle, emphasizing that mistakes should be explained and communicated understandably to the users. Finally, the tenth heuristic describes the need for proper documentation of tasks. This way, the user can follow concrete next steps or search for further instructions.

Together, Nielsen's ten heuristics (1994) represent the theoretical lens through which to analyse and assess Duolingo's usability and user experiences. As the analysis and discussion will show, many of the platform's strengths and limitations can be interpreted through the HCI principles.

Natural Language Processing

Natural language Processing (NLP) is a subfield of Artificial Intelligence (AI) that focuses on the relationship between computers and human language (Lan et al., 2024). It enables machines to understand, interpret, and generate human-like language. In its modern form, NLP can be tasked with speech recognition, information retrieval, and sentiment analysis. These capabilities are achieved through syntactic analysis (e.g., part-of-speech tagging), semantic interpretation (e.g., word embedding, semantic role labeling), and pragmatic understanding (contextual understanding) (Ashok Kumar et al.). In a broader context, NLP is pivotal in developing smart interactions between users and digital platforms. In an educational

context, NLP systems can analyze linguistic patterns and learner interactions to provide a personalized user experience, automated feedback, and adaptive content delivery.

The introduction of Large Language Models (LLMs) (Kamath et al., 2024) is a significant advance in NLP's development. These LLMs mark a considerable leap from the earlier task-specific models. LLMs are no longer trained to perform a single function; rather, they are pre-trained on a large corpus of existing textual data and designed to adapt to a wide range of language learning tasks through fine-tuning or prompt engineering. The core mechanism behind LLMs is the transformer architecture (Vaswani et al., 2017), which allows these models to handle long-range dependencies in texts and capture nuanced language patterns. The pre-training of these models involves predicting the next word in a text sequence; this way, the model learns grammar, factual knowledge, and gathers some reasoning capabilities (Kamath et al., 2024). This is done on commonly used datasets, network architectures, and learning objectives. General-purpose LLMs are trained on datasets, which are comprised of web-scraping, book corpora, and code. These data sources are further curated to standardize quality. The common network architecture refers to the already mentioned transformer architecture. It comprises an encoder and a decoder, which are leveraged for sequence-to-sequence tasks such as question answering. Finally, full language modeling (Kamath et al., 2024) is the most commonly used approach for the learning objectives. This approach involves the autoregressive prediction of the next token in a sequence given preceding tokens.

The following fine-tuning (Kamath et al., 2024) further adapts the model to specific use cases and reinforces learning from human feedback. The three types of recognized fine-tuning are instruction tuning, alignment tuning, and full-parameter tuning. Instruction tuning is based on the model learning to identify a variety of prompts. This technique involves compiling datasets with example tasks as natural instructions to make the model more responsive and adaptive to user prompts. Alignment tuning is conducted through reinforcement learning from human feedback loops (RLHF) (Kamath et al., 2024). This tuning is often undertaken to ensure that the model’s behavior is not only competent but appropriate. However, it is important to note that this technique involves subjective determinism on behalf of the annotators. Finally, full-parameter tuning refers to the traditional approach of updating all parameters of a pretrained model using a new dataset, tailored to a particular domain.

The principles of Natural Language Processing (NLP) and Large Language Models (LLMs) present the theoretical framework for this study. This lens highlights how personalization and pedagogical effectiveness are negotiated through computational design. As such, this theoretical standpoint will be used to interpret users' personalization learning experiences and examine how they perceive the relationship between AI and language learning.

Methodology

This section presents the methodological approach used to determine how users perceive personalization, behavioural design, and privacy. It introduces the research psychology, the design and implementation of data collection methods, and the analytical strategies employed to interpret this study's findings.

Philosophy of science

The philosophy of science is primarily concerned with two key aspects of research: ontology and epistemology (Saunders et al., 2019). The ontology standpoint is engaged with the nature of reality, it defines the assumptions the researcher would adopt about the world around them. This view can be further categorized into subjective and objective perspectives. Subjectivism holds the understanding that reality is socially constructed through the perceptions and actions of individuals. Objectivism, conversely, considers that social phenomena exist independently of human perceptions and actions. In essence, these views discuss how the researcher perceives the world and positions their study within it (Saunders et al., 2019). The research of this study stands to position itself in the subjective formation. This study acknowledges that reality is not objectively fixed but is constructed through individual experiences and interactions. This research focused on individuals' perspectives on personalization and privacy, adopting the belief that every user constructs their own reality through interactions and experiences with the platform. If the researcher were to take an objectivist standpoint, then they would have to observe and measure users in a detached manner, assuming there is a single reality to be constructed. This would divert attention away from the meanings of individual experiences, which would not aid in understanding how users perceive the trade-offs between personalization and privacy.

Epistemology concerns itself with the nature of knowledge and what is considered acceptable knowledge (Saunders et al., 2019). This view can be divided into three perspectives:

positivism, interpretivism, and realism. Positivism views reality as objective, where acceptable knowledge is derived from observable facts and evidence. This suggests that the researcher needs to adopt a detached role in the world in which the study is situated. Interpretivism connects to subjective reality and views knowledge as socially constructed. This approach enables the researcher to highlight the significance of lived experiences and individual interpretations. The third perspective, realism, strikes a balance between the two previously outlined views. It acknowledges the existence of an external reality, while still accepting the idea that social and cultural contexts shape it. With the three perspectives presented, this study took an interpretative stance, as its main aim is to understand how users make sense of their personal experiences with personalization and privacy.

This study adopted a pragmatic approach to its methodology. Pragmatism enables a more nuanced, outcome-driven approach, focusing on the practical applications of empirical findings (Kelemen & Rumens, 2008). The researcher believed that no singular philosophical position presents the whole truth about the world around us. Therefore, a mixed-methods approach is required to address the questions explored in this study. These questions relate to understanding users' experiences with personalization and privacy on the language learning platform Duolingo and how these insights can be applied in practical settings.

Research design

The process of conducting this study can be divided into several key steps. First, data about the platform Duolingo is gathered and thoroughly examined to filter out the most relevant information. Second, the underlying theories connected to digital platforms, Human-Computer Interaction, and Natural Language Processing are chosen. This part adopted a primarily descriptive approach (Saunders et al., 2019), as its primary focus is to contextualize Duolingo and the relevant theories in the world in which the thesis is situated. A mixed-method approach is adopted based on the knowledge collected from these sources. Open- and closed-ended questions were formed to gauge how users perceive personalization, their experience with Duolingo, and the degree of privacy they are willing to share with the company. This part of the research is primarily exploratory in nature, as it seeks to contribute to the existing body of knowledge on the subject (Saunders et al., 2019). The findings from the survey were then further discussed on the basis of the theoretical framework to attempt to determine what practical insights can be derived from them. The final part was deductive

(Saunders et al., 2019), as it tests how the chosen theoretical concepts resonate with real-world user data; however, it is framed by a pragmatic approach. Thus, this study explored several approaches, developing a multifaceted methodology that incorporates descriptive contextualization, exploratory investigation, and deductive interpretation, in an effort to achieve real-world insights and practical relevance. Furthermore, Generative AI (Chat GPT-4) was used in this study as a tool for language refinement and structure suggestions to ensure a cohesive presentation of ideas.

Survey design

As this study aimed to collect a diverse view of how users perceive personalization and privacy, the online survey is the most suitable option for primary data collection. The survey was hosted via Google Forms, where participants were greeted with a brief introduction to the purpose of the following questionnaire, which further informed them that their answers would be kept anonymous. To ensure a higher response rate, the survey was designed to be brief, with a completion period between 5 to 7 minutes. The complete survey can be seen in Appendix B.

The survey was available in English, ensuring consistent accessibility and comprehension of the content provided to the participants. Both closed and open-ended questions are included, however, the survey was predominantly composed of closed-ended questions. They consist of multiple-choice answers, binary (yes or no) responses, and likability scales, designed to collect measurable data about users' motivations, preferences, and comfort level. The open-ended questions were intended as follow-ups to previous closed-ended ones, where participants were given the opportunity to express different motivations, concerns, or critiques of the platform. Combined, the two sets of questions provided statistically measurable data about attitudes and preferences, while giving insights into the subjective meaning each user attaches to their experience.

The survey was structured into three main sections: first, it starts by collecting information about general user behaviour (frequency of use, primary motivation for using the app, etc.). Secondly, it addressed questions related to engagement with personalized features, including lesson adaptivity, motivational features, and notifications. And finally, it aimed to investigate how comfortable users are with sharing various types of personal information, including

location, daily activity, and typing patterns. The survey's segmentation was grounded in HCI principles (Nielsen, 1994), which emphasize the importance of understanding user experiences across various stages of interaction with the platform. The survey focuses on usability, user trust, cognitive load, and feedback.

Data collection

Saunders et al. (2019) argue that surveys are a particularly effective way to collect data for exploratory purposes. The choice of conducting a survey was constructed of a few key components: the limited amount of time, the number of researchers, and the aim for standardized answers. A key benefit of a web-based survey is its ability to reach a bigger, geographically dispersed pool of participants. This type of data collection enables multiple people to complete the questionnaire simultaneously, making the process more efficient. As the researcher of this paper is only one, taking them out of the equation of actually conducting interviews themselves saves time. It enabled more efficient data collection across a broader sample. Cross-sectional time sampling (Saunders et al., 2019) was chosen due to the limited time available to the researcher during the thesis writing period. The participants, once provided with the link, had a timeframe of a week to complete it. The survey, in total, collected 59 answers, all of which were completed and are be used for the research.

The survey was distributed in the researcher's academic program group through a social media platform. That approach utilizes a convenience sampling method (Saunders et al., 2019), as these participants presumably had previous experience using the app and were most likely willing to participate in the study. As the academic program of the researcher is conducted in English, a second language for most participants, this sample group is likely to consist of users who are or have used Duolingo. Furthermore, the sample group comprised individuals in their twenties, a demographic group that plays a significant role in adopting and using emerging technologies.

In addition to the primary data collected from the survey, this study also used secondary data from Duolingo's official website. This includes publicly available information about the platform's features, user interface, use of Artificial Intelligence, and data privacy policies. The secondary data provided understanding of the platform's structural and functional

architecture. This data also allowed the study to identify and explore gaps between Duolinogo's official claims and the lived user experience.

Data analysis approach

Due to the mixed-methods approach chosen for data collection, qualitative and quantitative techniques were employed in the data analysis. This approach also aided the pragmatic orientation of the study, as it enabled the exploration of both measurable data and user insights. A descriptive statistics approach (Saunders et al., 2019) was employed to analyze the closed-ended questions from the survey (quantitative data). Frequency distribution and cross-tabulation uncovered patterns in user engagement, personalization perception, and data privacy concerns. Pie and bar charts were used to assist in interpreting these findings. These visualizations provided a more comprehensive overview of trends and aided in presenting comparisons between different respondent groups (current vs. past users, short-term vs. long-term). In addition, to analyse the open-ended questions a thematic coding approach (Boyatzis, 1998) was employed. This approach allowed the researcher to identify patterns, themes, and key findings. The researcher followed Boyatzis' (1998) method of first familiarizing with the data, then generating codes that reflect meaningful segments of the given answers. Furthermore, these codes were clustered into general themes, carefully reviewed to ensure coherence and distinction, and further presented and analyzed in the following analysis section.

Limitations

While the methodology standpoint of this study is shaped to balance practical and analytical content, its limitations should also be acknowledged. Firstly, convenience sampling (Saunders et al., 2019) restricts the generalization of the empirical findings. While it may offer a general perspective on how university students perceive personalization and privacy in language learning, it does not constitute a universal guidebook for Duolingo. Furthermore, the choice of cross-sectional time sampling (Saunders et al., 2019) is time-limited and does not allow for the collection of continuously evolving user behavior over a more extended period. If the researcher had a broader team and more time, examining a more geographically and age-dispersed sample group over a longer period would have benefited the study's practical findings.

Secondly, although the study employs a mixed-methods approach, the quantitative, closed-ended questions predominate, which skews the balance inherent in this approach's nature. The reduced qualitative data limits the depth of the findings and may lack the nuance that the study would have gained from follow-up interactions with the participants, where they could clarify and expand on their answers. Additionally, the researcher is choosing to do a single case study. As described by Yin (2004), a method called 'triangulation' should be employed to develop such a study. This method outlines the importance of using three different methods for data collection and rejects the idea of trusting a singular source. This study acknowledges the importance of this method; however, as time and resource constraints were in play, the researcher chose only to use the online survey and Duolingo's official statements.

Despite the outlined limitations, the methodology approach remains appropriate for the study's exploratory aim. The findings provide a starting point for understanding how users in the age of digital language learning apps perceive privacy and personalization.

Results and their meanings

This next section of the paper presents the survey results and highlights the insights that can be drawn from them. Given the mixed-methods approach employed in the analysis, this section is divided into two parts. First, the quantitative results and their meanings are presented, followed by the qualitative findings.

Quantitative analysis

The presentation of the quantitative findings is conducted in the following outline:

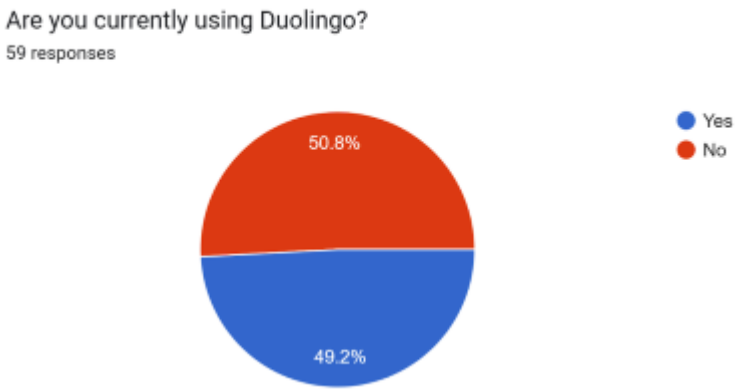
1. Findings on User Engagement
2. Findings on User Perceptions of Personalization and Motivation
3. Findings on Data Collection and Privacy Awareness

User engagement

The survey respondents present an almost perfect split between those currently using the app (49.2%) and former app users (50.8%) (*Fig. 1).* This creates an opportunity to gain balanced insights into both user retention and user churn. From the distribution of usage frequency

[Appendix A, Figure A1], it becomes clear that 30 (50.8%) of the participants are or have used Duolingo every day. From these daily users, 19 are current, and 11 are former app users. The same figure shows that 13 (22%) responders use the app infrequently, classified as 'once a month'. Of them, 11 are past app users, which suggests that their disinterest in it may be why they stopped using it. The third majority comes from 9 users (15.3%) who use the app 'Every second day'. The majority (5 participants) are current app users, indicating their interest in the app. The total number of active participants who use Duolingo as often as every day to every second day is 24. This suggests that a significant number of respondents (40.6%) have incorporated Duolingo into their daily routines, presenting a compelling sample of individuals who enjoy using the app.

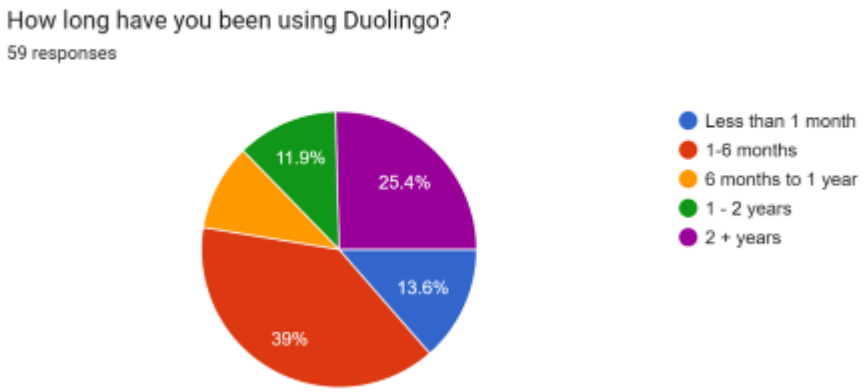
Fig.1: Are you currently using Duolingo



The daily usage of the responders [Appendix A, Figure A2], again, appears to be an almost ideal split between the two categories. 54.2% (32) of the users report that on the days they use the app, they spend less than 15 minutes doing so. Eighteen of those users are not current app users, suggesting that their lack of commitment and perhaps interest are the reasons for their limited interactions with the app. While 14 of the current users report using the app for less than 15 minutes a day, this raises the assumption that their time might be limited or their commitment might not be substantial. Furthermore, 44.1% (26) of the respondents report using the app for between 15 and 30 minutes on the days they have. They are further split into 15 current and 11 past app users, showing a strong commitment from both types of users. From these insights, it becomes clear that mid-level usage results in longer user retention than lighter daily usage.

Lastly, for this part of the analysis, it is xamined how long the responders have been using the app (*Fig.2*). The split this time is quite diverse: 39% of people report that they have been using Duolingo for '1-6 months', while the second largest group, at 25.4%, has been using it for '2+ years'. With these two majorities, the study is formed on the basis of a balanced distribution between newcomers and long-term users. 39% of responders are 23 people, of whom 18 are not current app users, showcasing a significantly high drop-off rate in a short period (1-6 months). Among the responders who have used the app for over two years, 12 out of 15 still use it. This indicates that Duolingo excels in retaining long-term users. Of the responders (13.6%) who reported using Duolingo for 'less than a month', none are current users anymore. While 6 out of 7 responders, who have been using the app for '1-2 years', are current users. And finally, 5 out of 6 responders in the '6 months to 1 year' category are currently using the app. These findings present an interesting contrast between Duolingo's inability to create an engaging early-stage experience and its strong long-term user retention over an extended period of time.

Fig.2: How long have you been using Duolingo for?

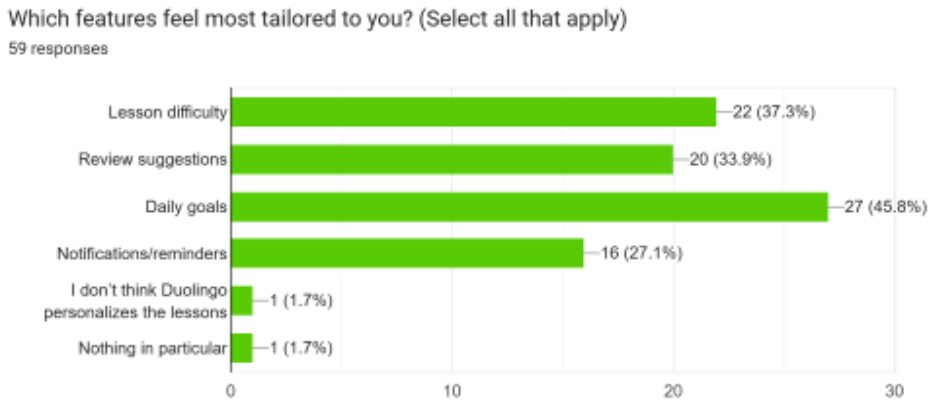


Perception of personalization

The following subsection examines the findings regarding users' perceptions of personalization and their motivations. From the *User perceived degree of personalization in Duolingo* [Appendix A, Figure A3], it becomes apparent that more than 50% (52.5% - 31) of respondents believe that Duolingo personalizes their experience, although only 'slightly'. The remaining respondents are distributed between 'moderately', at 16.9%, and 'very', at 5.1%,

with ‘not at all’ accounting for 25.4%. Of the last, 8 out of 15 users are not current, which leads to the assumption that the level of personalization is a relatively significant reason for user churn. To further support this claim, 15 out of the 31 who believe the app adapts slightly are also no longer using it. To further investigate the degree of personalization the participants feel, they are asked a more specific question about how they evaluate the degree to which Duolingo adapts to their strengths and weaknesses [Appendix A, Figure A4]. The weak feeling of personalization is boosted further, as 55.9% of users respond with ‘somewhat’. Only two responders (3.4%) confidently answer ‘yes’, while 14 responders (23.7%) confidently answer with ‘no’. Considering that there is a significantly higher number of users who are unsure if they have experienced personalization (10 responders - 16.9%), compared to those who can confidently say they have, this also supports the claim that Duolingo’s personalization mechanisms are not well-perceived by users.

Fig.3: Features that feel the most tailored to individual learners



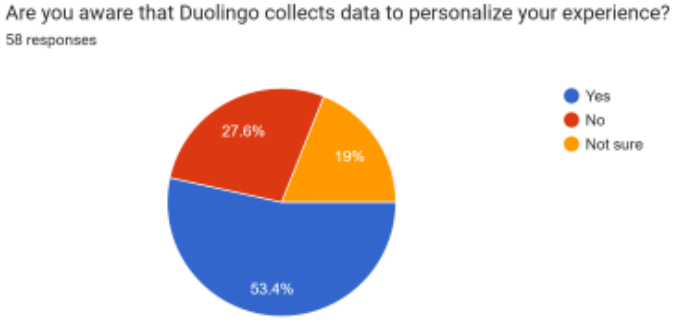
To delve deeper into the perceived personalization, the survey asks the respondents to identify which aspects of their learning process they found most personalized. They are given a multiple-choice option between 4 preselected values: lesson difficulty, review suggestions, daily goals, and notifications/reminders. Furthermore, they are also given the option to add a value themselves. From *Fig. 3*, it becomes evident that daily goals and lesson difficulty take the leading positions, with 27 answers (45.8%) and 22 answers (37%), respectively. Review suggestions take third place with 20 answers (33.3%), followed by notifications/reminders with 16 answers (27.1%). The most common combinations of answers are: daily goals and lesson difficulty (6), review suggestions and lesson difficulty (5), and review suggestions, lesson difficulty, and notifications/reminders (7). Between the participants, only two took the

opportunity to express personal feedback; however, they did not knowingly become a notable minority, as both expressed that nothing from their experience felt personalized to them. This question is specifically designed to follow the previous one, as the intention is to delve deeper into personalization, However, the findings from it raise a conflict as it becomes evident that when respondents are presented with specific values to choose from, a very stong majority of them (57 of 59) notice a form of personalization. While in the previously asked question, 24 respondents noted that they either do not acknowledge personalization or are unsure of it.

As outlined in the theoretical framework, this paper is strongly shaped by Human-Computer Interaction (HCI) theory, particularly by the concept of gamification. Which is why it is essential to ask the respondents about their perception of the motivational gamified design of the app. The users are again given a multiple-choice option between preselected values: XP points, Daily Streaks, Leaderboards, Badges/Achievements, and Gems. Again, they are also given the option to express personal feedback. From the *Most motivating Duolingo feature* [Appendix A, Figure A5], it can be seen that Daily Streaks stand out as the most motivating feature, selected by 44 (74.8%) participants. Second place is shared by XP points and Badges/Achievements, with 18 answers (30.5%), followed by Leaderboards with 16 answers (27.1%). There is a noticeable difference between the first and second place, highlighting the effectiveness of streak-based incentives (habitual app use). The little interest expressed by responders in gems or currency-like rewards suggests that rewards are less likely to sustain user engagement. Only two users again express their thoughts, one of which points out again that noting seems to motivate them. However, the overall perception of the gamified features appears to be positive, which creates an interesting dynamic in contrast to the negative perception of personalization.

Data collection and privacy concerns

Fig.4: Data collection awareness



This subsection presents and analyzes the findings on users' perspectives on Duolingo's data collection practices. First and foremost, it was essential to determine whether users are aware that the app collects their data. As indicated in *Fig.4,* the majority (31 respondents, 53.4%) are knowledgeable about the data being collected. However, a significant portion, 27.6% (16 respondents), is unaware of this data collection, while the remaining 19% (11 people) are unsure. Considering that this creates an almost even split between knowledgeable and unknowledgeable respondents, concerns about Duolingo's transparency and information sharing arise. These findings highlight a significant disparity between the app and its users regarding trust-building and informed consent.

To investigate the level of transparency users feel, they are asked whether they consider Duolingo transparent in its use of user data. Although a vast majority of users are aware of data collection, almost half, 49.2% (29 respondents), have never considered Duolingo's level of transparency. Less than 9% (8.5%) believe that Duolingo is transparent in its data policies, while 15.3% firmly say it is not. The remaining 25.4% have noted that the app is somewhat transparent [Appendix A, Figure A6]. The findings from these two questions reveal a key tension: while users are mostly aware that their data is collected, they lack the knowledge or confidence in how Duolingo communicates its policies transparently. This conclusion further suggests two possibilities: either users are not concerned by the data collection, thus do not engage with the available information, or Duolingo does not provide a clear and accessible way to their data collection and usage policies.

Furthermore, the participants are asked a series of questions in which they had to rate their comfort level with different types of data Duolingo collects on a scale of 1 to 5, ranging from least to most comfortable. *Fig. 5* provides an overview of the findings from these questions. Users are most comfortable sharing their learning progress and the time of day they use the app. With 27 responders giving a 5 and 23 a 4 to the learning process tracking, respondents demonstrate that they are at ease with the app tracking metrics, which directly correlate with its performance and functionality. Time of day use is a relatively comfortable metric, as the ratings are mainly distributed between 4 (22 answers) and 5 (19 answers). However, the comfort levels drop regarding typing patterns, with the most common answer being 3 (21 responders). This reflects a more neutral stance rather than a negative one, considering that, in total, between 4 and 5, there are 26 responders, while between 1 and 2, there are only 12.

Fig.5: Comfort levels with data collection types

Data collected by Duolingo / Level of comfort	1	2	3	4	5
Learning progress	2	2	5	23	27
Time of day use	1	4	13	22	19
Typing patterns	4	8	21	12	14
Speech/audio recordings	16	14	12	7	10
Location data	29	18	6	2	4

In the last two rows, the app's most sensible type of data collection is stated, and it also reflects a noticeable shift in user comfort levels. Regarding speech and audio recordings, it is evident that the answers are distributed between 1 and 3, with a significant number of users at a comfort level of 1 (16 individuals). However, there is no considerable gap between the number of users in the three rankings, as comfort level 2 is marked by 14 individuals and comfort level 3 by 12. Although skewed toward the lower comfort levels, this relative balance suggests a more moderate level of doubt or uncertainty among users. In contrast, the level of discomfort is significantly more pronounced when it comes to collecting location data. A substantial majority (29 respondents) rated their comfort level as the lowest (level 1), and 18 selected level 2. In total, 47 out of the 59 participants expressed a considerably low level of comfort with sharing their location data. In contrast, only six users express a high level of comfort, with ratings ranging from 4 to 5. This significant discomfort suggests that users perceive location data collection as invading their personal space. As seen from these findings, users react more positively to data being collected when it could be beneficial for developing a better learning experience for them, rather than for Duolingo's statistical use.

To conclude, the survey asks if users would be willing to share more personal data in exchange for improved personalization [Appendix A, Figure A7]. Most respondents, 61% (36 individuals), reflect their reluctance to share additional personal information with the app. Only 13.6% (8 people) said that they would be willing to share. However, a notable number of users (15-25.4%) indicated "depends" as their answer. The users who chose "depends" are asked to express their conditions for sharing more information. Their insights are analyzed in the next section, as they are relevant to the qualitative data analysis.

Qualitative analysis

This section presents and analyzes the qualitative data collected through the survey. The open-ended questions explored in this section are developed to help users express their experiences freely beyond predetermined options. The responses are coded and organized into thematic areas:

1. Users' Motivations and Goals
2. Frustrations and Disengagement
3. Data Collection Concerns

Fig.6: Thematic analysis of open ended survey responses

Theme	Codes	Quotes
Motivation for use	Learning for fun Career growth Travel Connection Casual use	"To connect with people and learn for fun." "It is fun to use in my free time to help me freshen up my knowledge." "It distracts me from everyday tasks." "To advance in my career"
Disengagement	Loss of motivation Ineffective experience Notification fatigue	"I didn't find it effective." "It felt like the same thing over and over again. I just got bored at some point." "I got annoyed by the notifications."
Time constraints	Busy schedule Low priority	"I was too busy." "I had phases of using it, but when I got busy. I dropped it."
Privacy trade-offs	Depends on benefits Transparency Scepticisms	"What types of personalization they would do" "Depends on what data" "Benefits and how useful it would be for me" "If I could separate speaking from all other types of lessons"
Scepticism about personalization	Lack of felt impact Vague benefits	"I can't feel any personalization right now with the data gathered."
Data sensitivity	Unwillingness to share certain data types	"If it's too personal and I don't see the correlation to helping me learn languages." "I don't see the need for Duolingo knowing my location"

User motivation and goals

From a practical perspective, knowing the motivation behind users' engagement with the app is essential. This way, the study will provide real-life insights into user retention patterns and help Duolingo tailor its design to better align with user goals.

The responders are asked to define their primary goals for using the app. While the quantitative data shows that 64.4% of users engage in "learning for fun" [Appendix A, Figure 9], the open responses heavily backed up these numbers. The users who decided to share their personal experience mainly describe the app as a "casual tool" or a "fun distraction". Among the answers, some people point out that it helps them connect with a more diversified group of individuals or to prepare for their travels. Only one participant expressed that he uses the app to help him advance in his career. These answers combined lead to the idea that the participants perceive the app as more of a fun, low-stakes side activity rather than a serious educational platform. These findings raise questions about the platform's identity and whether users prefer it from a convenience standpoint rather than believing in its ability to be a comprehensive solution for language fluency. While Duolingo has succeeded in framing language learning as enjoyable, this quality may unintentionally diminish its educational value.

Frustrations and Disengagement

A significant portion of the qualitative data collected centers around the reasons users had to stop or pause using the app. This focus is intentional, as understanding friction points and factors behind disengagement is crucial for guiding future improvements. The main themes that emerge from the findings are: loss of motivation, ineffectiveness of the process, annoyance at notifications, and lack of time. The quantitative data shows that the most commonly cited reasons are the loss of motivation and time constraints. The qualitative findings mostly back up these arguments. One participant says: *"It felt like the same thing over and over again. I just got bored at some point."* (Fig.6). Another points out, *"I didn't find it effective"* (Fig.6). This shows that Duolingo's purposeful design of forming routines risks becoming too repetitive, leading to disengagement, particularly if the user does not see learning gains.

Vital feedback from the participants also indicates their frustrations with using the app. Duolingo's notification systems seem too pushy for users, as some noted that they get annoyed by them. This can heavily contribute to their decision to drop off the app, particularly if they don't see it as a way to acquire proper educational knowledge. Additionally, many users reported that time constraints are a problem. Although the app is designed to be a time-efficient tool, users still deprioritize it when juggling their day-to-day responsibilities. This is perfectly illustrated by this user's quote: "*I had phases of using it, but when I got busy, I dropped it."* (Fig.6). From these findings, it becomes clear that most users perceive Duolingo as a casual, low-commitment tool used for fun. For Duolingo to stop simply retaining occasional users and genuinely engaging long-term ones, it must address its challenges with repetitiveness, notification fatigue, and lack of learning depth.

Data collection concerns

The final open-ended question asks users to elaborate on their willingness to share more personal data in exchange for better personalization. As the quantitative data shows, a significant portion of the users expressed a conditional desire to share more information [Appendix A, Figure A7]. The central themes that emerge from their conditions are: transparency, benefits, and data type. Respondents seem aligned on the idea that further data sharing is only acceptable if they see notable benefits. Whether that would be better personalization or tailored experiences, most are united in that they need to get something valuable in exchange for their personal information. However, with that said, some respondents outline that it depends not only on the benefits but also on the type of data the app would request to collect. As one noted, *"If it's too personal and I don't see the correlation to helping me learn languages,"* (Fig.6), they would not feel comfortable sharing information they do not deem educationally oriented. Another participant expresses the concern: *"I don't see the need for Duolingo knowing my location"* (Fig.6), showing a hint of distrust regarding the platform's current data collection practices. Further backing up the argument, if people do not see the correlation between the collected information and their active learning process, they are less likely to continue using the app.

One participant also expressed doubt in Duolingo's ability to translate the collected personal information into a personalized experience: *"I can't feel any personalization right now with the data gathered."*(Fig.6). This statement opens up two potential lines of interpretation: This

user will be more likely to share personal data in the hope of better personalization, or they will be more skeptical about sharing data, as they have never seen its benefits. In the context of personalization, a user expressed their desire to be able to *"separate the speaking from all the other types of lessons."*(Fig.6). This indicates that users would like to be in control of the modularity of lessons, leading back to the importance of the benefits derived from the data collected.

Discussion

The foregoing analysis illustrated the complex interplay between Duolingo's algorithmic personalization mechanics, its gamified motivational approach, and users' opinions and feelings toward data sharing. The following section of this study unpacks the key tensions that arise when examining this interplay through the established theoretical lens. The discussion is be divided into four sections:

1. Adaptive Feedback and Systems Visibility
2. User Control and Customization
3. Motivational Mechanics vs. Learning Coherence
4. Data Transparency, Privacy, and Trust.

Adaptive feedback and System Visibility

A critical tension emerging from the mixed-method analysis is the disconnection between Duolingo's adaptivity features and the user's subjective experience with personalization. Quantitatively, 97% of responders recognize specific personalization features (daily goals, lesson difficulty, or review suggestions) when prompted with explicit options [Appendix A, Figure A5]. However, when asked if they feel that their experience is personalized, only 52% experience 'slight' personalization, while a substantial minority is unsure or receives no personalization at all [Appendix A, Figure A3]. This creates the traction between perceived vs. actual personalization, signaling Duolingo's failure to convey system responsiveness clearly to users. Nielsen (1994) outlines this struggle as his first heuristic: visibility of system status, where he emphasizes the importance of keeping the users informed about what is happening. Contextualizing it, this principle demands that the platform not only adapt its

content based on the scenes but also clearly communicate how and why it has adapted. The current shortfall is showcased in the survey's open-ended questions, which reveal that many learners are unaware when the system has tailored their experience. It is echoed through qualitative responses like, *"I can't feel any personalization right now with the data gathered."* (Fig.6).

Consequently, the lack of perceived personalization risks undermining users' motivation and trust in the platform. As shown in the analysis, users do not clearly identify personalization and do not perceive that the app knows and learns from their past experiences, thus creating the belief that their efforts do not yield meaningful results. Moreover, as Burston (2014) has outlined in his research on pedagogical challenges, clearly communicated feedback is the backbone of a cohesive learning trajectory. While users are left feeling like their efforts are not taken into account, there should be no doubt as to why they would feel demotivated or not express interest in the learning process at all. To counter this tension, Duolingo should consider presenting users with more explicit design cues as to why their adaptations change the way they do. Presenting them with progress cues such as progress bars, weak-word weekly streaks, or lesson previews allows them to mentally map how the app analyzes their previous experience and adapts their future steps.

Optimizing adaptive feedback is not only a way to polish the interface but also directly correlates with user motivation and long-term retention. When the learners can see why they are practicing certain materials, they will start perceiving the platform as an effective tutor rather than a repetitive game. This might significantly improve the user's retention rate. As seen in the analysis, the early stage dropouts are mainly users who struggle to see knowledge acquisition, describing their learning process as the same thing all over again (Fig. 6). These users will benefit greatly from a clear explanation of their learning journey. Presenting them with explanations like: "Why this lesson? - You missed 4 out of 10 past-tense verbs in the last section, so we have curated these exercises to help reinforce that skill." or "Weak-word weekly streaks - You have misspelled the following words multiple times this week (example words). Practice them now to turn your streak green!". Ultimately, this will create the feel that each exercise is purposeful and strengthens the learning journey the user is on.

Enhancing adaptivity and visibility will be a strategic investment from Duolingo. It will clarify the value of interactions, thus deepening users' trust in the learning process. Further, it

will reinforce the platform's pedagogical credibility, therefore becoming an app that users actually utilize for learning purposes rather than a superficial habit-former.

User Control and Customization

In the context of personalization, this section explores the recurring friction point between Duolingo's standardized lesson progression and users' desire for individualized learning journeys. The app's minimalistic design and instructional flow are created in such a way as to reduce cognitive overload and aid the learner efficiently through the process. However, that same unified design offers little genuine control. Regardless of their previous objectives or experiences, every user must follow the same vocabulary, grammar, and speaking exercises structure. Currently, the app does not give users control over how their lessons are organized or the ability to prioritize specific skills or modules. This one-size-fits-all approach forces the learners into the same learning path, causing frustration among users with diversified learning needs. In the qualitative data, this message is echoed through comments like: *"If I could separate speaking from other types of lessons",* or *"It felt like the same thing over and over again"* (Fig.6). This outlines the concept that users are unhappy about the rigid sequencing, diminishing their sense of agency.

From a theoretical standpoint, Nielsen's heuristics (1994) represent a structured lens through which these shortcomings can be examined and potentially redeemed. According to the principle of flexibility and efficiency of use (Nielsen, 1994), interfaces should adapt to users' styles and preferences in learning to better accommodate both novices and experts. Duolingo's current linear lesson model prevents more experienced users from moving past the foundationally structured units, undermining efficiency. At the same time, the repetitive module lessons could be perceived as demotivating to novice users, who quickly grow bored of repeating identical tasks. To address this issue, the platform should utilize the user control and freedom heuristic (Nielsen, 1994). With the correct mechanics in place, the system could offer users the function to override the constructed lesson flow and allow them to structure their content according to their timely needs. Embedding a customized session toggle would enable users to swap tasks around, such as choosing between listening sessions, grammar practice, or speaking, so that every session aligns better with their goals. Thus, it creates a sense of freedom in the learner, positioning them as the primary driver of their success.

Duolingo is not unfamiliar with the mechanics behind such innovations, as it would require a simple optimization of its established NLP and LLM infrastructure. By further training its large-language models on datasets of user interactions and learning pathways curated from its millions of users rather than relying on broadly scraped public datasets, the platform will be able to unlock far more precise pedagogical personalization. Utilizing the trained models, the app can move forward from simply serving fixed sets of sentences that somewhat fit the learning trajectory of the user, to providing specifically created, coherent content (grammar, listening, or speaking exercises) that reinforce case-specific vocabulary the learner has struggled with.

Another content reorchestration would allow the platform to organize its content based on specific exercise sets rather than solely on thematic groups. This reorganization will enable users to enter particular practice sequences based on their immediate needs. Additionally, if trained and tuned further, their in-house LLMs can become responsive to free-form user requests and map them out to the appropriate content created for them. The model would be able to recognise a user requesting context-specific content, then retrieve the corresponding grammar exercises from the Duolingo datasets and create a cohesive sequence of tasks containing contextualized exercises and immediate feedback.

Ultimately, focusing on further model training and tuning will allow Duolingo to empower its users with control over their learning trajectories. Giving this control over to the learners will deepen trust and show that Duolingo actively adapts to their needs and wants, rather than trying to sell the one-size-fits-all model.

Motivational Mechanics vs. Learning Coherence

Another central paradox that stems from the analysis is that Duolingo's gamified experience, while highly effective at driving user engagement, falls short in translating this into coherent language learning. From the quantitative data, it is known that 74.6% of the users identify Daily Streaks as their primary motivator for keeping up their progress [Appendix A, Figure A8]. In comparison, gamified elements like XP points and badges are less motivating, with close to 30% of users identifying them [Appendix A, Figure A8]. This significant gap between the first and second motivator, and the first motivator being Daily Streaks, leads to the assumption that users are motivated by features that form habitual use. However, at the

same time, from the qualitative data (Fig.6), it has become clear that the most common reasons for users to give up on the app are the loss of motivation and their perceived ineffectiveness of the learning process. This is where the friction is created, as motivational features succeed in fostering daily returns but fail to deliver convincing evidence for real learning gains. Consequently, users find themselves caught in a cycle of maintaining streaks without experiencing learning progress, therefore undermining Duolingo's place as an effective language tutor from their perspective.

This cycle perfectly highlights one of the frictions Burston points out in his critique of MALL's pedagogical challenges. He identifies the risk of micro-learning activities fostering only surface-level progress rather than meaningful comprehension, which proves to be the case among Duolingo users. In the context of this study, responses like *"It felt like the same thing over and over again"* (Fig.6) and the substantial number of respondents who report that they don't find the process effective illustrate Burton's friction. In other words, Duolingo's gamified tasks may trigger the " challenge-success-reward" loop (Zichermann & Cunningham, 2011), however they remain hollow unless genuine language learning is triggered. Duolingo's bite-sized tasks are another feature that contributes to creating this tension. Theoretically, they are designed in such a way as to reduce cognitive load through microlearning. However, in practice, this reduction of interactions into isolated words or sentence-level exercises does not consistently build transferrable real-life skills. With how few users note that they use the app to advance in their education or for travel purposes (education - 11.9% and travel - 8.5%) [Appendix A, Figure A9], it becomes evident that respondents struggle with developing practical skills like starting conversations, comprehensive listening, or writing.

However, an interesting contradiction arises when another aspect of the quantitative data is examined. A notable group (39%) of responders disengages within the first six months of using the app (Fig. 2). Meanwhile, users who reach the two-year mark show higher retention (12 of 15 are still active users). Two explanations emerge from the presence of this data: long-term users have recognized real value from the learning journey they have been on, or they have continued purely out of habit, maintaining their streak almost mechanically. Which explanation holds true for the distinctive user likely depends on their personal goals and threshold for perceiving incremental progress. However, combined with the findings discussed in the previous paragraphs, the long-term retention seems to be guided mainly by

habit formation. The split between 39% of users having used the app for up to 6 months, and 26% using it for more than two years (Fig. 2), with a minority of users in the other time categories, speaks clearly of habit formation at the 6-month benchmark. This pattern underscores that Duolingo primarily relies on its gamified routines to cement habit rather than foster coherent learning milestones.

A promising strategy for reconciling Duolingo's habit formation lies in its implementation of advanced NLP, more specifically LLMs. Utilizing LLMs' ability to generate contextually meaningful tasks that mirror real-world language use could be the solution to the traditional micro-learning of decontextualized words or phrases. However, this likely solution brings back the discussion from the previous sections about the perceived vs actual personalization. Without communicating its personalization choices, Duolingo risks falling prey to the same invisibility that undermines its existing adaptability. The key findings from this part of the discussion tie together an important yet overlooked insight: habit alone cannot sustain meaningful language learning. While habit-forming mechanisms like Daily Streaks or XP points are perfectly engaging for return visits, yet they remain surface-level unless accompanied by knowledge acquisition. Additionally, coherent learning journeys require each exercise to be presented and positioned as a deliberate step in a pedagogical journey, which would be impossible without personalization.

To this end, transparency around how LLMs tailor users' experiences becomes critical for transforming mechanical habits into meaningful language learning processes. This transparency will ensure long-term user engagement reflects real skill development rather than mere routine.

Data Transparency, Privacy, and Trust

The final but perhaps most delicate tension deducted from the analysis is the clash between users' data collection concerns and the data collection needed for personalization. When users are asked to identify which data Duolingo collects is unsettling, they identify location data, speech, and audio recordings as the most problematic (Fig. 5). The qualitative analysis backs up this theme, as multiple users mention their struggle connecting the relevance between their location and learning journey (Fig. 6). From the table for *comfort levels with data collection types* (Fig.5), the most unproblematic data for users is the learning progress and time of day

usage. This correlates with the qualitative analysis, as the majority of responders have outlined that they would be willing to share personal data solely if they see the correlation with their learning progress. Typing patterns seem to be in the gray area, neither comfortable nor uncomfortable. However, the overall feeling toward sharing more personal information is negative, as 61% of responders refuse to share additional information with the app [Appendix A, Figure A7]. The minority of 25%, which have indicated dependents, further express clear conditions like data transparency and demonstrable learning benefits.

The reluctance to share more personal data creates a problematic feedback loop: users expect personalization in return for data collection, but when that personalization is absent or unclear, their willingness to share diminishes. However, the reduced data limits the platform's ability to personalize effectively, thus undermining the adaptive learning progress promised to users. This then becomes a vicious cycle, further diminishing users' trust in the platform's usefulness.

Turning back to Nielsen's heuristic about system status visibility (1994), this issue becomes an extension of the problems with actual vs. perceived personalization. Just as users need cues to understand how lessons have adapted to their learning style, they also need transparent communication about why data is collected and how it benefits them. When users sign up to use the Duolingo app, they are greeted with the terms of service and privacy policy, which indicate what data is collected and how it is used. However, in the digital world in which people nowadays live, they rarely put in the effort to read through these policies, as it is common knowledge that refusal equals foreclosing access to the service. Consequently, informed consent is rarely achieved, meaning that Duolingo's transparency around its data practices is formally in place but remains invisible to the average user.

To address this issue, Duolingo could employ a progressive disclosure model. Instead of asking about permissions all at once, they could request them individually during the course of the learning process. For example, when the learner reaches the speaking exercise, alongside asking them if the app could use their microphone (currently a standard request), it could also provide them with an explanation as to why their voice data is recorded and how it might be used if they agree to share it. An explanation like: "May we use your voice recording from the speaking exercise, when tailoring pronunciation feedback to help you with accent nuances?" leaves no room for questioning why the app would collect this information.

Furthermore, as it becomes evident from the survey, users are very sceptical about sharing their location data (Fig.5). However, if they were prompted with the explanation of: “Enable location services to receive idiomatic expressions from your region”, would the skepticism remain the same? Alternatively, even simpler, “We collect location data for internal statistics to determine our biggest markets or the retention rate of certain countries. ”. Each explanation should clearly indicate the educational value users would get out of sharing that specific data. This way, Duolingo would be able to turn privacy concerns into a source of empowerment for users, making them the primary informed drivers of personalization.

To this end, the four sections of the discussion seem undeniably interconnected and lead back to the central conflict of mistrust in the platforms' approaches due to the little informative feedback users receive. Collectively, these insights point out that Duolingo's next step would not be further investment in sophisticated algorithms but rather a comprehensive redesign of how these algorithms are surfaced, controlled, and justified to users. Making personalization visible and user-driven should be the app’s main priority in its quest for building user trust and forming sustained, meaningful language learning.

Reflections and Suggestions for Future Research

The following section of this paper is designed to critically reflect on the findings and shortcomings of this study's framework, analysis, and discussion sections. Furthermore, it will present areas for improvement, suggest alternative approaches, and include considerations for future research.

Reflections on the study

While previous sections have evaluated users’ interactions and perceptions of Duolingo, it is also important to review the choices and constraints that shaped this study. Starting from the initial framework formation, Nielsen’s heuristics provide a coherent foundation for evaluating Duolingo. However, an alternative theoretical lens could have provided clearer insights into user motivation and habit formation. Furthermore, by employing theories revolving around the socio-cultural aspect, the study could have gained insights into the community formation Duolingo advocates for through their competitions and leaderboards. Such insights would have been beneficial for creating a broader picture as to how users manage their

self-motivations and whether communities influence collaborative learning behaviours. Combining theoretical frameworks on user motivations with the established focus on Nielsen (1994) and Burston (2014) could have yielded a more nuanced taxonomy of motivational factors. Thus, presenting Duolingo with key items they need to either redesign or reinforce.

Having mentioned the existing choice of framework, Nielsen's ten heuristics (Nielsen, 1994), and Natural Language Processing (Ashok Kumar et al.), it is important to highlight their vast age gap. While Nielsen's beliefs were developed in the 1990s, a relatively static age, NLP and LLMs operate in the dynamic data-driven world nowadays. This could impose meaningful implications for the practical insights of this study, as the nature of the theories is rooted in contradicting assumptions. While Nielsen's heuristics propose a stable, consistent menu-driven interface, NLP and LLM frameworks operate on continuously evolving user data. Consequently, aligning the heuristics' imposed predictability and consistency with the dynamically changing structure of NLP/LLM could prove to be difficult.

Furthermore, ethical considerations when conducting this study should be mentioned. In the survey context, all participants were explicitly informed that their responses would remain anonymous and that their data would only be used to inform this study. Therefore, consent was obtained under the principle of informed anonymity. This decision was made to create an environment where participants could freely express their opinions without being influenced by social bias.

Finally, the researcher themself is an essential aspect of the choice of topic for this case study. As a past Duolingo user, I recognize that my personal familiarity with the particular platform could have introduced selection bias. While Duolingo is a clear market leader in Mobile-Assisted Language Learning, my previous experience with the app is a key reason for the choice of the final case company. To mitigate the possible biases from my own perspective and interactions with the app, the study was designed to include both qualitative and quantitative data, ensuring a balanced portrayal of user experiences, not relying solely on my own individual impressions.

Future Research Suggestions

Several future research opportunities have emerged from the ground provided by the discussion section. Firstly, it is essential to highlight that the study's findings have been

collected over a one-week time frame. For a deeper dive into users' interactions with Duolingo, an outgoing longitudinal design approach (Saunders et al., 2019) could be employed. With findings from such research, it will become apparent whether early-stage dropouts correspond to unmet personalization expectations and whether long-term users derive educational gains versus practice out of mere habit.

The potential solutions provided in the discussion section could also be a further research topic. While, in theory, explaining data collection to users in a timely manner sounds like a reasonable idea, proper testing should be conducted to determine if that would change users' attitudes. Such a study would be able to decide whether transparency truly mitigates privacy concerns or simply adds to the cognitive load that apps like Duolingo are meant to reduce. Similarly, a case study could be conducted on feedback visibility and customization toggle suggestions. Thus, it would be able to answer the question of whether this new prototype would improve user satisfaction and trust or cause unwanted cognitive overload.

Conclusion

This study has situated Duolingo within the framework of digital platforms, which have transformed the traditional linear value creation into dynamic ecosystems, where value is created through user feedback and algorithmic adaptation. Duolingo is a prime example of how the educational industry could benefit from this new system, where users are both value consumers and creators. Due to users' dual role in the emerging framework, it is critical to examine their subjective interactions with the platform to gain a comprehensive understanding of how design, algorithms, and data practices shape learning outcomes.

By combining the practical insights from the mixed-method approach with the foundational framework for Human-Computer Interactions and Natural Language Processing, this study has outlined four main tension points that shape the Duolingo experience:

1. Adaptive Feedback and System Visibility
2. User Control and Customization
3. Motivational Mechanics vs. Learning Coherence
4. Data Transparency, Privacy, and Trust

The first friction point surfaces the gap between perceived and actual personalization, which stems from the lack of proper explanation to users. The second point highlights users' frustrations with the app's one-size-fits-all approach when designing lesson flows, which fails to accommodate individual goals. Thirdly, the motivational mechanics vs. learning coherence shows how habit-forming gamification drives frequent returns, but falls short in translating this into genuine learning gains. Finally, the fourth point uncovers a feedback loop in which unjustified data collection triggers unwillingness to share, thus undermining the very personalization it is meant to provide.

Taken together, these findings reveal that Duolingo's future growth does not depend on the expansion of algorithmic complexity but instead on the reconfiguration of human-centered design, which surfaces, controls, and justifies algorithms use to learners. By making personalization visible and explainable, Duolingo can turn its habit-forming mechanisms into meaningful language acquisition and empower users to be active architects of their own educational journey.

Bibliography

Ashok Kumar, L., Karthika Renuka, D., & Geetha, S. (Eds.). (2023). Deep Learning Research Applications for Natural Language Processing. IGI Global.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage Publications.

Burston, J. (2014). MALL: the pedagogical challenges. Computer Assisted Language Learning, 27(4), 344–357.

Duolingo - Learn a language for free @duolingo. (n.d.). Duolingo. Retrieved April 2, 2025, from: https://www.duolingo.com/approach

Duolingo - Learn a language for free @duolingo. (n.d.-b). Duolingo. Retrieved April 2, 2025, from: https://www.duolingo.com/super

Duolingo Press room. (n.d.-c). Retrieved April 2, 2025, from: https://press.duolingo.com/#about

Duolingo - Research. (n.d.-d). Retrieved April 3, 2025, from: https://research.duolingo.com/

Henry, P. (2025, February 26). *How Duolingo uses AI to create lessons faster*. Duolingo Blog. Retrieved April 3, 2025, from: https://blog.duolingo.com/large-language-model-duolingo-lessons/

Information Resources Management Association, editor, & Association. (2022). *Research anthology on developments in gamification and game-based learning* (Association, Ed.). IGI Global.

Kamath, U., Keenan, K., Somers, G., & Sorenson, S. (2024). *Large Language Models: a deep dive*.

Kelemen, M., & Rumens, N. (2008). *An introduction to critical management research*.

Kukulska-Hulme, A., & Shield, L. (2008). Overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. ReCALL, 20(3), 271- 289

Lan, Y., Li, X., Du, H., Lu, X., Gao, M., Qian, W., & Zhou, A. (2024, January 15). *Survey of Natural Language Processing for Education: Taxonomy, Systematic Review, and Future Trends*. arXiv.org.

MacKenzie, I. S. (2012). *Human-Computer Interaction: an Empirical Research perspective*.

Nielsen, J. (1994). *Usability Engineering*. Academic Press

Palalas, A. (2011). Mobile-assisted language learning: Designing for your students. In S. Thouesny & L. Bradley (Eds.), Second language teaching and learning with technology: Views of emergent researchers (pp. 71–94). Dublin

Parker, G. G. , Choudary, S. P., & Van Alstyne, M. W., (2016). *Platform Revolution.* W.W. Norton & Company Ltd.

Saunders, M., Lewis, P., & Thornhill, A. (2019). Research methods for business students, 8th ed. In *Pearson eBooks*.

Schumpeter, J. A. (1942). *Capitalism, socialism, and democracy*. Harper & Brothers.

Sharples, M., Taylor, J., & Vavoula, G. (2005, October). Towards a theory of mobile learning. Mlearn 2005 Conference Proceedings. Cape Town,

Topal, M., & Karaca, O. (2021). Chapter 1: Gamification in e-learning., In *Research anthology on developments in gamification and game-based learning* (pp. 1–20). IGI

Global.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All You Need. *arXiv (Cornell University)*, *30*, 5998–6008.

Werbach, K., & Hunter, D. (2012). For the win: How game thinking can revolutionize your business. Wharton Digital Press.

Yin, R.K. (2004). Case Study Research: Design and Methods. 3rd Edition, Chongqing University Press, Chongqing.

Zichermann, G., & Cunningham, C. (2011). Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps. Sebastopol, CA: O'Reilly Media.

Appendix

Appendix A: Supplementary Survey Findings

Figure A1. Distribution of respondents' Duolingo usage frequency

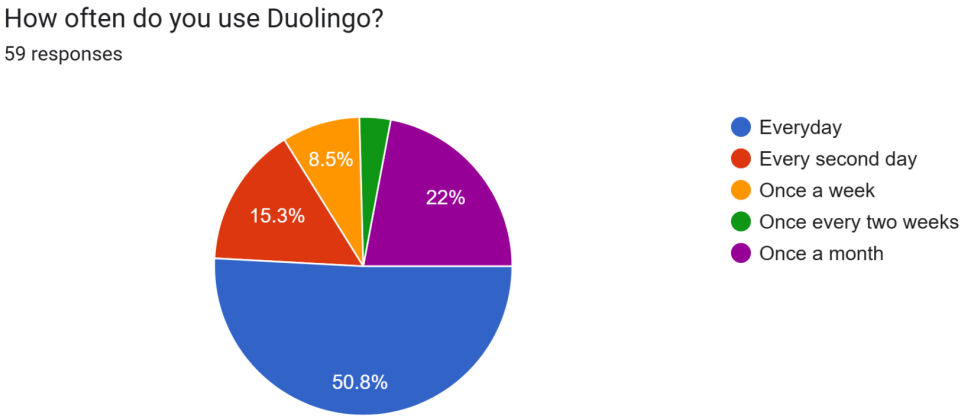


Figure A2. Daily usage of respondents

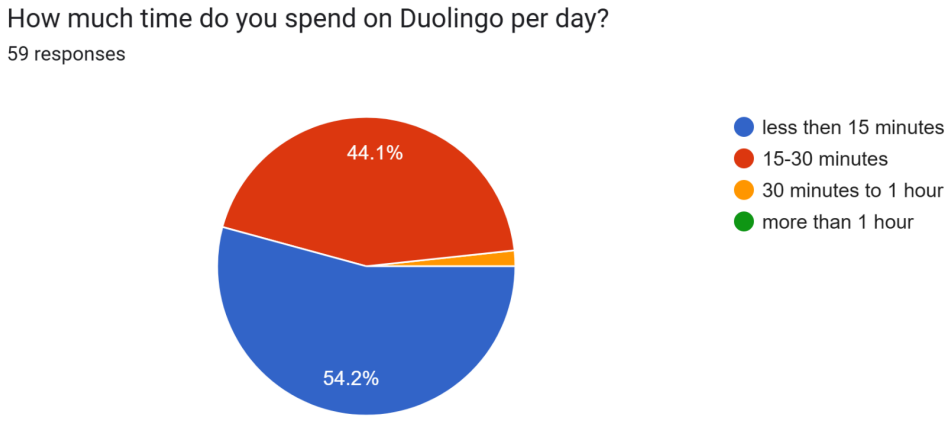


Figure A3. User-perceived degree of personalization in Duolingo

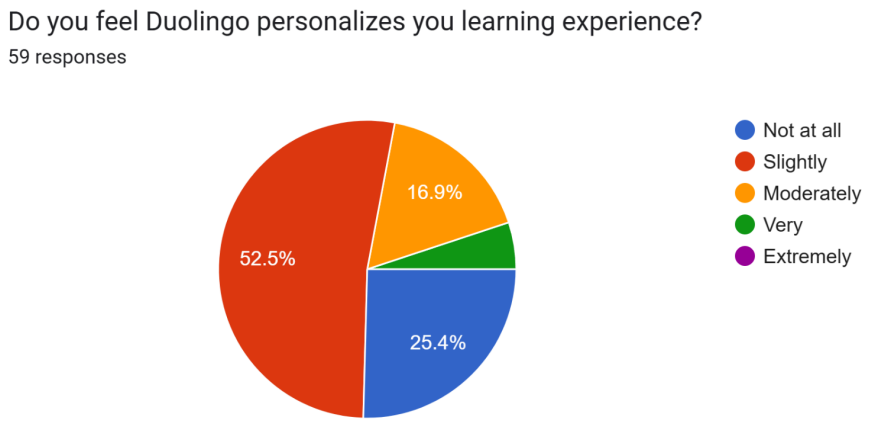


Figure A4. User perception of Duolingo's adaptation to individual strengths and weaknesses

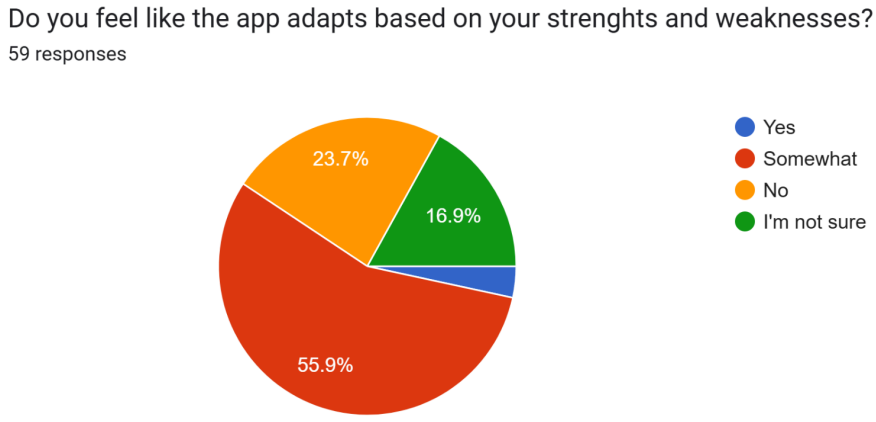


Figure A5. Most motivating Duolingo features

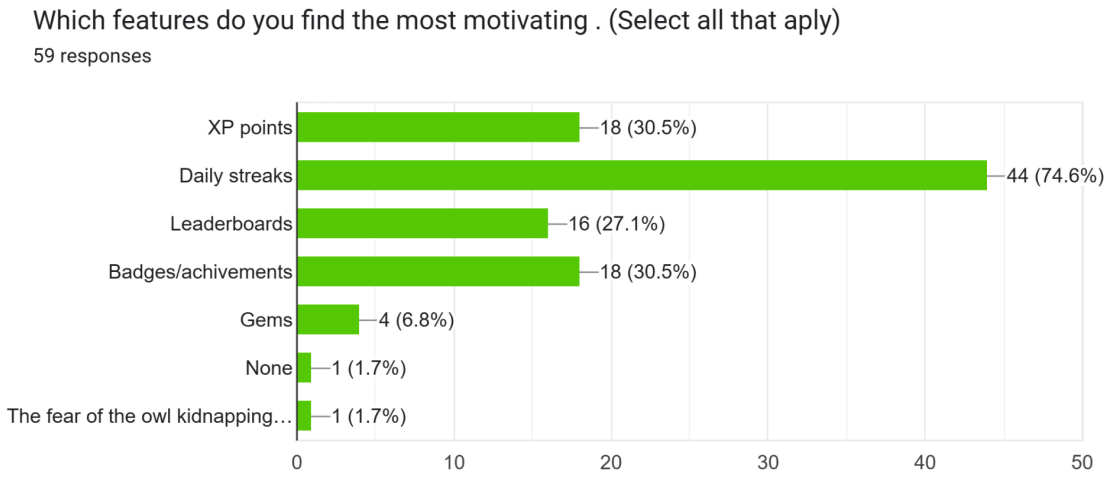


Figure A6. Perceived transparency of Duolingo's data practices

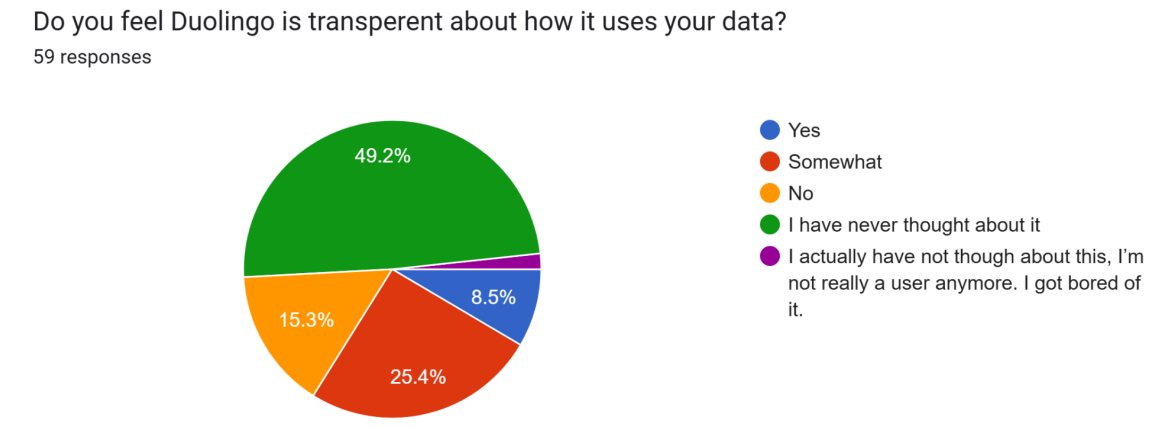


Figure A7. Users' willingness to share additional personal information

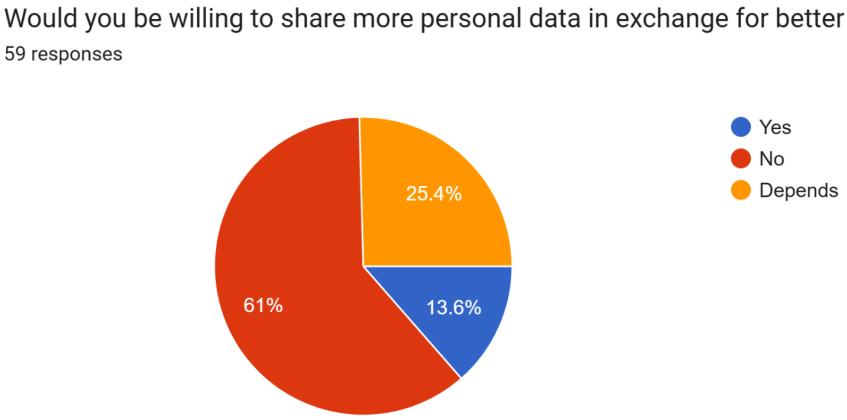


Figure A8. Users' perception of feature personalization

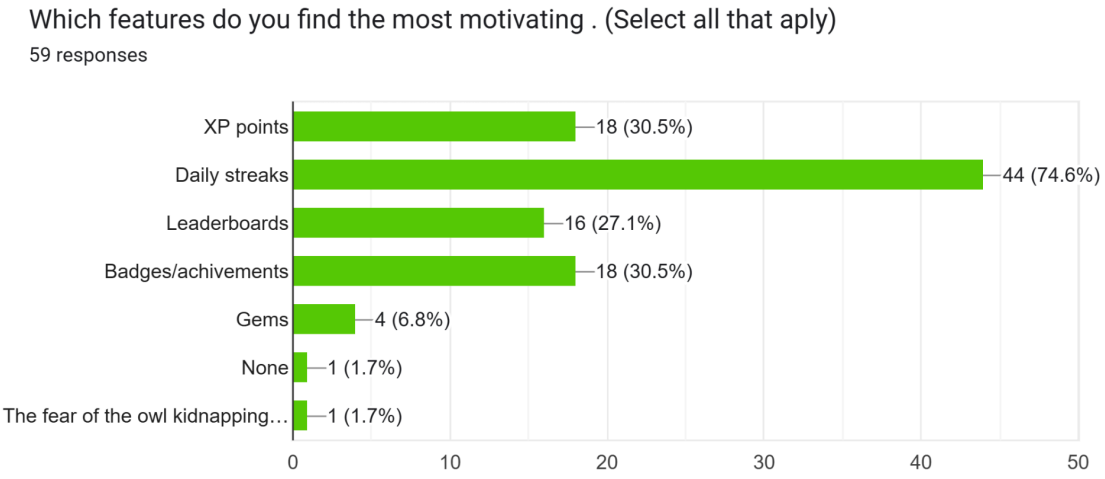
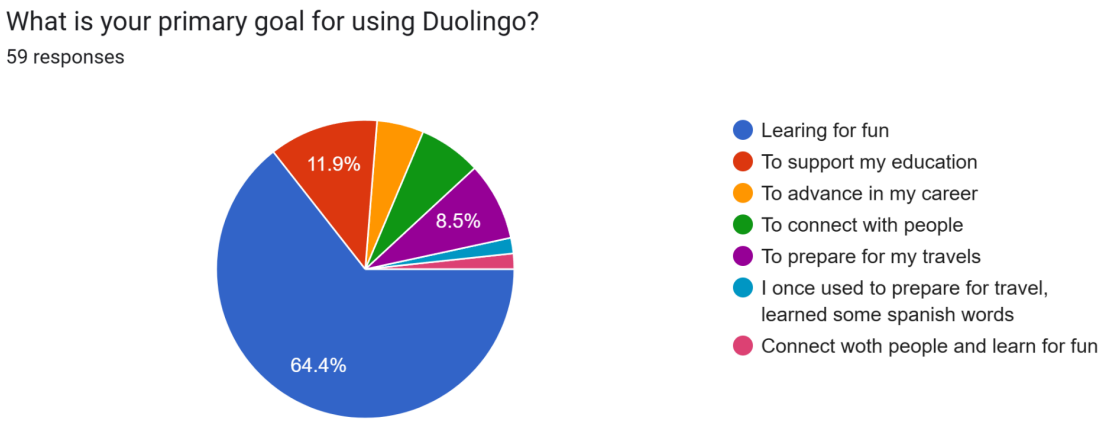
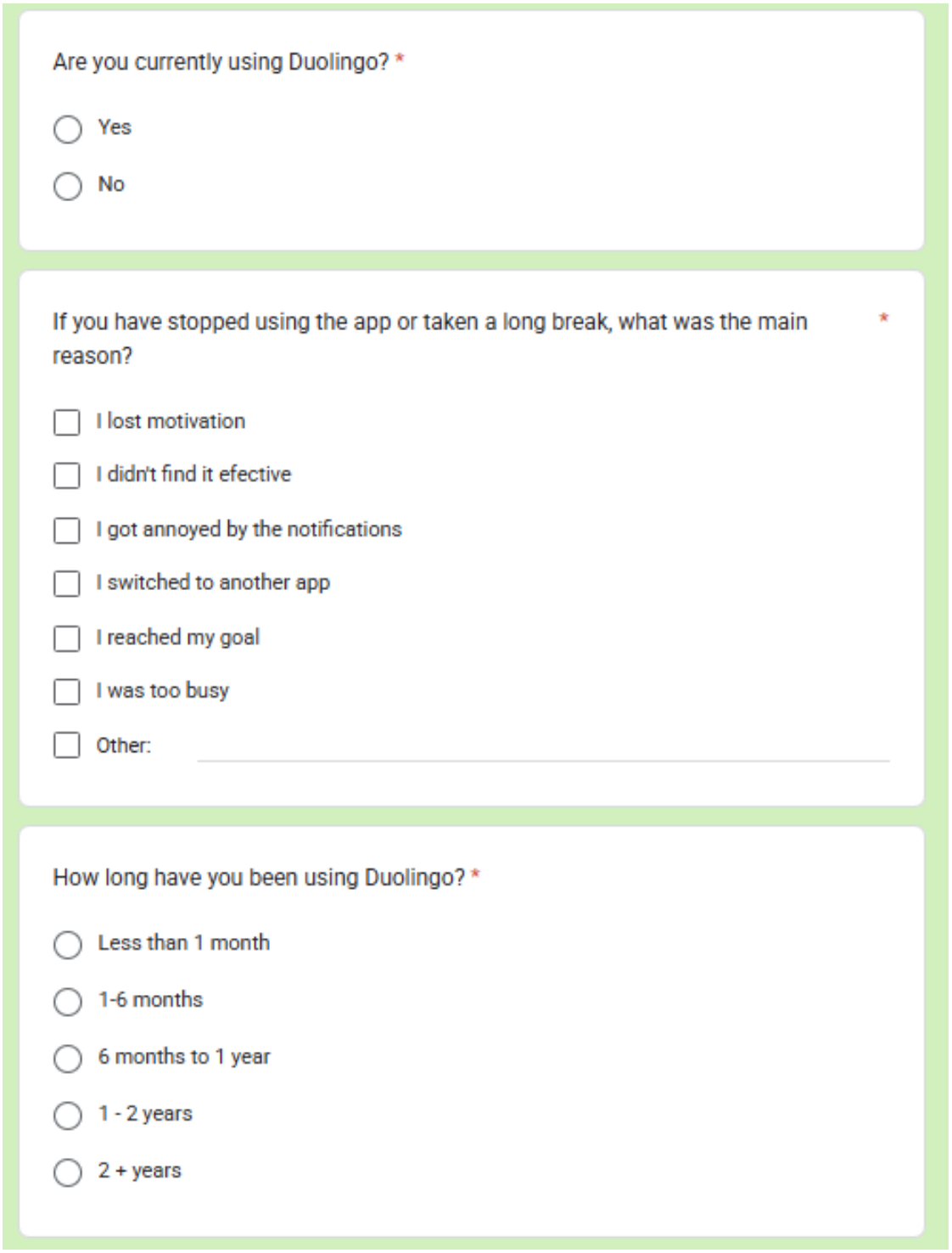


Figure A9. Users' primary goals for using Duolingo



Appendix B: Survey about the Use of Duolingo

https://forms.gle/QtZyG6riyv788jW88 - *Link to the Google Forms Survey*



Are you currently using Duolingo? *

- ◯ Yes
- ◯ No

If you have stopped using the app or taken a long break, what was the main reason? *

- ☐ I lost motivation
- ☐ I didn't find it efective
- ☐ I got annoyed by the notifications
- ☐ I switched to another app
- ☐ I reached my goal
- ☐ I was too busy
- ☐ Other: _____

How long have you been using Duolingo? *

- ◯ Less than 1 month
- ◯ 1-6 months
- ◯ 6 months to 1 year
- ◯ 1 - 2 years
- ◯ 2 + years

How often do you use Duolingo? *

- ◯ Everyday
- ◯ Every second day
- ◯ Once a week
- ◯ Once every two weeks
- ◯ Once a month

How much time do you spend on Duolingo per day? *

- ◯ less then 15 minutes
- ◯ 15-30 minutes
- ◯ 30 minutes to 1 hour
- ◯ more than 1 hour

What is your primary goal for using Duolingo? *

- ◯ Learing for fun
- ◯ To support my education
- ◯ To advance in my career
- ◯ To connect with people
- ◯ To prepare for my travels
- ◯ Other: _____

What language(s) are you learning on Duolingo? *

Your answer

Do you feel Duolingo personalizes you learning experience? *

- ◯ Not at all
- ◯ Slightly
- ◯ Moderately
- ◯ Very
- ◯ Extremely

Which features feel most tailored to you? (*Select all that apply*) *

- ☐ Lesson difficulty
- ☐ Review suggestions
- ☐ Daily goals
- ☐ Notifications/reminders
- ☐ Other:

Do you find Duolingo's notifications usefull?

- ◯ They keep me on track
- ◯ They are a nice reminder
- ◯ I mostly ignore them
- ◯ They are too frequent
- ◯ I find them annoying

Do you feel like the app adapts based on your strenghts and weaknesses? *

- ◯ Yes
- ◯ Somewhat
- ◯ No
- ◯ I'm not sure

Which features do you find the most motivating . (*Select all that aply*) *

- ☐ XP points
- ☐ Daily streaks
- ☐ Leaderboards
- ☐ Badges/achivements
- ☐ Gems
- ☐ Other: _____

Have you used features of Duolingo Max such as Roleplay, or Explain my Answer? *

- ◯ Yes, regularly
- ◯ Yes, a few times
- ◯ No, i haven't tried them
- ◯ I don't know what those are

If you have used them, how helpful did you find the AI-driven feedbak?

	1	2	3	4	5	
Not helpful	◯	◯	◯	◯	◯	Extremely helpful

Are you aware that Duolingo collects data to personalize your experience?

- ◯ Yes
- ◯ No
- ◯ Not sure

How comfortable are you with Duolingo collecting the following personal data?

Learning progress *

	1	2	3	4	5	
Not comfortable at all	◯	◯	◯	◯	◯	Extremely comfortable

Time of day usage *

	1	2	3	4	5	
Not comfortable at all	◯	◯	◯	◯	◯	Extremely comfortable

Typing patterns *

	1	2	3	4	5	
Not comfortable at all	◯	◯	◯	◯	◯	Extremely comfortable

Speech/audio recordings *

	1	2	3	4	5	
Not comfortable at all	◯	◯	◯	◯	◯	Extremely comfortable

Location data *

	1	2	3	4	5	
Not comfortable at all	○	○	○	○	○	Extremely comfotable

Would you be willing to share more personal data in exchange for better personalization? *

- ○ Yes
- ○ No
- ○ Depends

Please specify what does it depend on.

Your answer

Do you feel Duolingo is transperent about how it uses your data? *

- ○ Yes
- ○ Somewhat
- ○ No
- ○ I have never thought about it
- ○ Other: _____