Research and Report
Duolingo: Like Language, We Never Stand Still.

220031985

The University of St. Andrews

Abstract. This case study is based on Duolingo, a rapidly growing online language-learning platform. The study specifies the system's development and maintenance from a software engineering standpoint, focusing on its functionalities and features that changed the field of its application domain, as well as its limitations in functionality and qualities. We also conclude with the ethical concerns that have arisen throughout its software lifecycle, as well as a brief discussion of possible future trends and system improvements that the platform could incorporate in future releases.

1 Introduction

Duolingo was conceived in Pittsburgh at the end of 2009 by Carnegie Mellon University professor Luis von Ahn and his post-graduate student Severin Hacker [1]. It is a web-based game-style language learning tool available as a mobile application [2]. The Duolingo Review 2022 says that with over 500 million users, it is the most popular language-learning platform and the most downloaded education app.

Duolingo offers a tree-based training strategy. It uses flashcards, TinyCards, images, listening, and writing lessons to motivate you to study unfamiliar words, phrases, and uncomplicated sentences [3]. Users learn by translating words and sentences from the target language into their native or well-known language and then eliciting the meaning of these words. As mentioned, an extensive tree skill-based network unlocks the progressing level which is based on a user completing the exercises to further progress down the tree and unlock higher levels/units. In terms of Language Learning, researchers have examined Duolingo to conclude that the learning platform is a promising supplementary tool for language learning but does not provide the learner with authentic language [2].

Our goal is to critically analyse, research, and report on Duolingo's software development, including its limitations in terms of functionality and quality,

how it plays a role in its application domain (education), its impact on the industry, and to conclude with a brief discussion on how it can be improved/developed further down the chain.

2 Software Development

2.1 Behind the Scenes

To successfully build a language platform included of a website and apps that can run on various operating systems, we must first understand the layers of technology that go into the construction. In 2011, Duolingo made use of the best services and solutions available, such as incorporating machine learning to aid in the development of its product. Duolingo runs on Amazon Web Services, SQL and DynamoDB to store data and use vocabulary. Python was used for the back end, while backbone.js, jQuery, and Twitter Bootstrap were used for the front end [4]. Though it has not been mentioned anywhere/clearly defined that Duolingo followed a specific type of software – development approach, I would side with the fact that they followed Rapid Software Development (Agile) based on the following findings:

- Duolingo was constantly evolving over the years implementing corrections on language courses and other bug fixes. [9,10] **(Embracing Change)**
- They needed to deliver software with new features and evolve it quickly (rapid software development) with the demand of the end-users. [8,11] **(Customer Involvement)**
- Constantly improving Duolingo's course creation process by creating simple tools to add more filters in the future and a metric that changes a course's quality [5] (**Maintain Simplicity)**
- The team implemented AI (Artificial Intelligence) to handle bottlenecks in course content and integrate it into features such as reminders [6,7] and built their machine-learning technology to personalize lessons [14]. **(People not process)**

2.2 Digital Game-Based Language Learning (DGBLL):

Mobile technology advancements have forever altered the way we learn and experience the learning process.

There appears a "hybrid mix of education and entertainment that weighs extensively on visual material, narrative or game-like formats and on more informal less didactic styles of address" defined as Edutainment which is less authoritarian and much more 'interactive' [12, p.282].

Duolingo is based on an adaptive (DGBLL) framework whose elements seem core to the education design are [2]:

1. **Multimodal:** (interaction factors, multimedia elements and narrative) sound effects on completing lessons, visual reference for new vocabulary words and audio for every word present in the sentence.
2. **Task:** (completion of challenges) completion of lessons, reaching the highest grade in each lesson or reaching the smallest criteria.
3. **Feedback:** (assessment of progress and rewards that are meant to simulate users' motivation) achievements, badges, gaining XP (experience points) and diamonds (in-app currency), the idea is creating positive-reinforcement conditioning.

2.3 Further Development & Sustenance

To keep up with trends and technical debt, applications like Duolingo had to learn, adapt, and evolve [4]. In addition to learning, the software had to run tests to ensure that users had an efficient learning experience and to figure out how quickly users learned languages. Future software builds based on feedback and testing from end users [9] called A/B Testing [16]. In its life cycle, Duolingo has gone through several modifications and changes, some of the major limitations and improvements are described below.

2.3.1 Rewriting Duolingo's Engine in Scala:

Duolingo had accumulated massive technical debt over the years, which needed rewriting code in a different codebase (i.e., Python to Scala) and halting any development of new features during this period, which could last several months.

Duolingo's rapid growth as a start-up meant that the Session Generator which existed from the start and had gone through all sorts of changes from Day 1 to resulting in the accumulation of technical debt.

Based on the detailed description given by Duolingo, we summarise the following findings [10]:

Images Illustrating before(L) and after(R) of Duolingo's Session Generator Architecture (Duolingo Blog: Rewriting Duolingo's Engine in Scala, Jan 31,2017)

Limitations:

- Every time the team needed a piece of data from a data store or cache an expensive computation, they created a dependency on a shared resource, resulting in the architecture shown here (L).
- The hard dependencies are highlighted in red, such as whether the Session Generator would fail if any of these or their respective networks failed (dependency). They stood for more network calls due to extensive networking within the dependencies, resulting in high latency rates.

Solution:

- To address this dependency issue, the Duolingo team created an architecture (R) that aimed to eliminate as many shared resources as possible, resulting in a simple and robust design.
- They were able to process course data offline and serialise it into files in AWS S3, Amazon Cloud Storage Service, which meant it only needed to fetch from a stable and inexpensive data store (Infrastructure as a Service, IAAS) and cache them rather than relying on other shared resources.

- The user data that offers a unique and personalised experience is fetched through API servers and injected into the request sent to Session Generator.

2.3.2 Redefining the Codebase:

Limitations:

The session generator was created in Python during the first stages of Duolingo development because it was simple to understand and useful to scuffle among developers from various backgrounds. That does not have to be the case because it has some disadvantages [10]:

- **Performance:** Python is slower than Java, C.
- **Memory Management:** Python is generally not thread-safe (leading to developers not utilising in-memory caches to their full potential).
- **Dynamic Typing:** Increase in runtime bugs due to the definition of interfaces between modules being weaker and the interpreter not being able to catch type-related bugs.

2.3.3 Moving to Scala

Aside from the new syntax, the team's first challenge when transitioning from Python to Scala was functional programming. The referential Transparency function produces the same output for the same input. It is one of the pillars of functional programming and goes hand in hand with immutability, which means you can't mutate state if you can't have side effects [10].

Scala proposes to solve some bottle-necks that other programming languages offer:

- **Less verbosity**
- **Less boilerplate**
- **Fewer bugs**

The highlights underlined by Duolingo were that [10] :

- **Latency had dropped by 98% from 750ms to 14ms**
- **Engine uptime increased from 99.9% to 100%**
- **Redesigned Architecture**
- **Refactored code from Python to Scala**

The rewrite was considered more stable, and the old infrastructure's degraded performance dropped from 0.1 to zero downtime in the first couple of months. As an added result, the number of bugs went down.

Limitations:

- Although it seems like a success, it is **far too early** to say if this redesign has **reduced the technical debt of Duolingo**.

3 Business Model

Duolingo, like any other company on the market, needed to generate revenue to support its resources and services, though this was not the case when it was officially launched as a start-up in 2009.

Though Duolingo is built on the freemium model (If you're not paying for it, you are the product), meaning that the core features of the product are free of charge to use [13] and if users wish to access the premium features such as unlimited hearts(unlimited mistakes), Gems/Ingots (in-app currency) and Duolingo Plus for non-ad interrupted service they can do so by opting to a subscription or buying the features.

Duolingo also offers Duolingo English Test, a certification test that gives the exam taker if passed, a certificate that proves his/her ability in English for $49 [13].

3.1 Duolingo Business Model Canvas

A Business Plan Canvas is a visual portrayal that serves as a strategic management template for developing new business models or documenting existing ones. **("Business model canvas and lean canvas: how to use in business planning")** It supplies an overview of the company's key partners, key activities, resources, and so on.

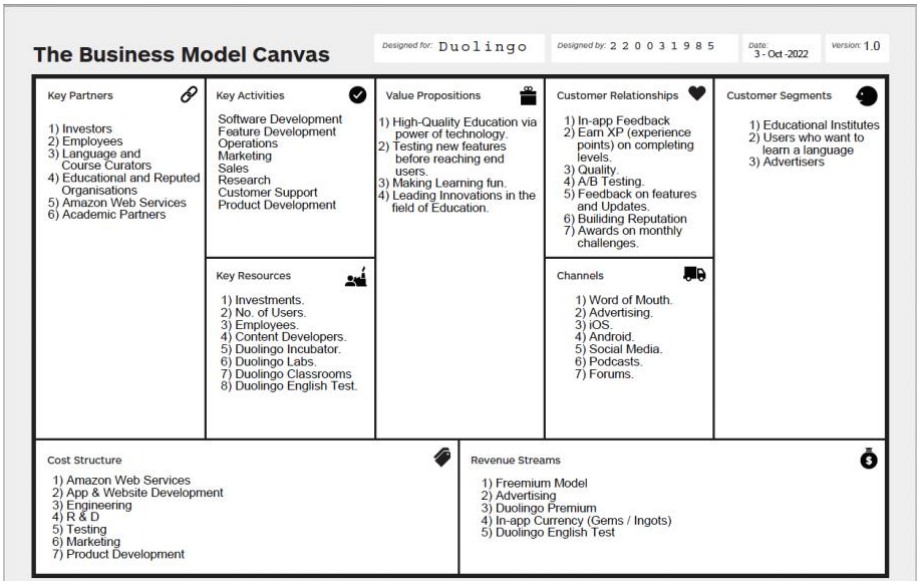

Fig: Business Model Canvas illustrating Duolingo's key assets [15]

4 Impact

"To bring high-quality education to everyone on the planet by harnessing the power of technology." – Duolingo

Duolingo's educational content and services have helped people all over the world, as the diversity of languages is a barrier to education and communication, as language barriers obstruct and impede education, cultural diversity, communication, and the spread of education [17].

Duolingo has managed to get over 500 million users around the globe and recent studies [18] dictate Duolingo curated courses in French and Spanish leading to comparable results as taught in a U.S University.

Innovations Duolingo developed that the team believes have caused a ripple in the education system are as follows [14]:

- Building a Business without charging educational content.
- Reinvented the wheel between brands and social platforms to connect with their audiences (e.g., Duolingo TikTok Channel has 4.8 million followers).
- Duolingo created a layer of management called "Duolingo for Schools" that sits on top of the Duolingo language learning app and allows teachers to create classrooms and invite their students by sending login information to the students. Teachers can create assignments and check students' classroom activities [19,20,21].

5 Ethical Concerns

Duolingo has managed to stay out of the spotlight in terms of ethical concerns, but some of its practises when viewed from a different angle, raise ethical concerns, such as:

- **Business Moral:** In the last decade, Duolingo relied heavily on "volunteers" from contributors, known as the Incubator Program, whose goal was to structure and curate language courses. While they were acknowledged for their participation, they were not compensated for their time in the initial stages. Though this was not the case when the programme began, Duolingo began to make money from the courses. It was forced to close the entire Contributor Community because the relationship could not be justified.

  The contributors were awarded the "Duolingo Language Impact Award," a $4 million fund shared among the community and various awards. The question is **whether they were properly compensated**, and **why close the "Contributor Program" when you could pay the contributors for their efforts.** I believe Duolingo took a giant step back and distanced itself from the language speakers and learners' community [22,23,24].
- Due to the app's sometimes **aggressive sense of humour** tries to ensure learners use it daily, Duolingo's mascot, a green cartoon named "Duo," has been scrutinised as an **internet meme** in which the mascot is "evil" and will stalk and threaten users if they do not continue using the app [25].
- The unmediated voices of 'ordinary' translation recipients contribute to the discursive (re-)construction of **translational ethics**. In one case, Duolingo, for example, supports a serious tone. Sometimes the threads hold comments on ethical issues, such as in the Esperanto thread, where one user complains about poor Esperanto translations where the publishers have either done a poor editing job (e.g., omitting properly accented characters) or else appear to run texts through Google Translate and publish them without post-editing [26].

- One could argue that Duolingo's use of "**crowd-sourcing**" **-** a practice of engaging a "crowd" or group for a common goal - often innovation, problem-solving, or efficiency [27] is frequently conflicted, as studies show that organisations and individuals face crowdsourcing challenges such as price undercutting, low pay, privacy, and deception, to name a few [28]. Luis Von Ahn, CEO, and co-founder of Duolingo is a well-known crowdsourcing pioneer, as his former company - ReCAPTCHA and Duolingo - show the utility of hidden crowdsourcing—that is, using behaviours that people already do to get work done. The beauty of this method is that it does not require any added conscious effort on the part of the people in the crowd to be effective [29].

6 Further Evaluation and Conclusion

Further evaluation and examination of Duolingo Services may reveal the following improvements:

- Ad frequency has increased to the point where it interrupts a user's Duolingo experience, causing them to uninstall the app and resulting in a loss of revenue and, more importantly, a customer.
- Duolingo's mobile application is limited functionalities when compared to its website and lacks some features such as discussion threads and duolingo schools.
- In terms of user-friendliness, the mistake system, represented by hearts and a decrease in count if mistakes are made by the user, may demotivate a user and ultimately lead to the app being uninstalled. The abolishment of such a system could lead to positive-reinforcement learning.

To summarise, Duolingo has overcome the challenges thrown at its services and has proven repeatedly that it is the frontier in the education field, breaking down language barriers through innovation and creating co-existential relationships between users and educators.

A user is lost when an app or service seemed to be unresponsive while that may be the case, Duolingo has proven to its users that it cares about the quality of service it provides by refactoring its services, although time and money were lost in the process.

When start-ups reach the big leagues, they commonly lose sight of their business goals, which are often dictated by their end users. However, by not charging for their educational content, Duolingo and its team have remained committed to providing a high-quality education to people all over the world and have managed to uphold the idea they began with**.**

Total Words: 2471

References

[1] Duolingo, https://www.studyfrenchspanish.com/duolingo-review, accessed: 02-10-2022

[2] Savvani, Stamatia (2019) State-of-the-Art Duolingo Features and Applications. In: International Conference on Interactive Collaborative Learning? -? Kos, Greece.

[3] Duolingo: About us, https://es.w3ki.com/wiki/Duolingo, accessed: 02-10-2022

[4] Software Development and Implementation on Language-Learning, https://www.rootstrap.com/blog/how-to-build-an-app-like-duolingo, accessed: 02-10-2022

[5] Improving Duolingo's Course Creation Process, https://blog.duolingo.com/how-were-improving-duolingos-course-creation-process, accessed: 02-10-2022

[6] AI use in course content, https://blog.duolingo.com/using-ai-to-open-up-bottlenecks-in-course-content-creation, accessed: 02-10-2022

[7] AI use in reminders, https://blog.duolingo.com/hi-its-duo-the-ai-behind-the-meme, accessed: 02-10-2022

[8] Migrating Android App from Java to Kotlin, https://blog.duolingo.com/migrating-duolingos-android-app-to-100-kotlin/, accessed: 02-10-2022

[9] Improving Duolingo, https://blog.duolingo.com/improving-duolingo-one-experiment-at-a-time/, accessed: 02-10-2022

[10] Rewriting Duolingo's Engine in Scala, https://blog.duolingo.com/rewriting-duolingos-engine-in-scala/, accessed: 02-10-2022

[11] Shifting from Objective C to Swift, https://blog.duolingo.com/real-world-swift, accessed: 02-10-2022

[12] David Buckingham, M.S.: Parental pedagogies: an analysis of British edutainment magazines for young children. Journal of Early Childhood Literacy 1(3), 281–299

[13] How does Duolingo make money, https://productmint.com/duolingo-business-model-how-does-duolingo-make-money/, accessed: 02-10-2022

[14] 10 Innovations that transformed education, https://blog.duolingo.com/duolingo-technology-innovations/, accessed: 05-10-2022

[15] Example of Duolingo Business Model Canvas, https://vizologi.com/business-strategy-canvas/duolingo-business-model-canvas/, accessed: 03-10-2022

[16] Duolingo - Using A/B testing to bring free language education to millions of users, https://youtu.be/xAx0ccvKw-M, accessed: 02-10-2022

[17] Language as a barrier to communication, http://encyclopedia.uia.org/en/problem/137300, accessed: 03-10-2022

[18] Duolingo Effective Learning, https://blog.duolingo.com/effective-intermediate-learning/, accessed: 04-10-2022

[19] Duolingo Learning Language Impact, https://www.impaakt.com/blog/impact-analysis-duo-lingo-language-learning-impact, accessed: 02-10-2022

[20] Blog on Duolingo for Schools, https://blog.duolingo.com/duolingo-for-schools/, accessed: 04-10-2022

[21] Duolingo for Schools Website, https://schools.duolingo.com, accessed: 04-10-2022

[22] Opinion: Duolingo Closing the incubator, https://www.reddit.com/r/duolingo/comments/vt78z9/opinion_duolingo_took_a_giant_leap_backwards_by/, accessed: 10-10-2022

[23] Ending and honouring our Course Contributor Program, https://www.reddit.com/r/duolingo/comments/m21f26/ending_and_honoring_our_course_contributor_program/, accessed: 10-10-2022

[24] Ending our Volunteer Contributor Program, https://blog.duolingo.com/ending-honoring-our-volunteer-contributor-program, accessed: 10-10-2022

[25] Duolingo's Wikipedia Page, https://en.wikipedia.org/wiki/Duolingo, accessed: 10-10-2022

[26] Annjo K. Greenall (2018) The discursive (re-)construction of translational ethics. 658-659

[27] Crowdsourcing, https://crowdsourcingweek.com/what-is-crowdsourcing/Crowdsourcing, accessed: 10-10-2022

[28] Ethics in Crowdsourcing, https://www.unomaha.edu/college-of-business-administration/center-for-collaboration-science/research/ethicsincrowdsourcing.php, accessed: 10-10-2022

[29] Working without even knowing it: ReCAPTCHA and Duolingo leverage unwitting crowds, https://d3.harvard.edu/platform-digit/submission/working-without-even-knowing-it-recaptcha-and-duolingo-leverage-unwitting-crowds, accessed: 10-10-2022