Second Brain

Read Once, Remember Forever

Team 64

Ria Sharma (rias@wharton.upenn.edu), Jake Brooks (jakebroo@wharton.upenn.edu),

Rohan Gupta (grohan@seas.upenn.edu), Vikram Singh (singhvik@seas.upenn.edu),

and Ricky Raup (rraup@sas.upenn.edu)

Advised by: Joseph Devietti and Henil Vipul Satra

Executive Summary

Second Brain is a multifaceted ecosystem and one stop shop for all information you have already consumed. We aim to facilitate your daily workflow, designed to streamline your research and knowledge management. We want to transform the way you interact with data, ensuring that every crucial insight from your past readings is just a query away. Join Second Brain and experience the power of having your own digital archive at your fingertips.

In terms of key features, we first provide users an easy interface to upload their PDF documents as well as online articles through a chrome extension. Second Brain stores your documents in folders that you design, but it also enables you to **automatically generate a folder hierarchy**, grouping together similar documents at the press of a button. We make it easy to **search** through everything you've read for the purpose of writing papers, familiarizing yourself with new research, or just exploring new interests. In addition to storing the content of your documents, Second Brain also **generates summaries** you can easily see by viewing your documents on our site. It further allows for you to chat with an LLM we host in order to **ask questions specifically related to your research** and fields you're interested in understanding better. In terms of interface, it's a file system that looks very similar to what you're already used to with Finder and Google Drive, but it makes it easier to search through your documents and upload information. It also has a chat feature to use our LLM.

In the following document, our team will outline our business plan, split into sections where we will outline our Unique Value Proposition, our Key Customer Segments, our Existing and Future Market Research, our Key Stakeholders, our Competition, and finally our Revenue and Cost model. Generally, these sections follow the business model canvas and should paint a strong image of the company we plan on building out here.

Value Proposition and Key Features:

Second Brain's core value proposition is the following: We enable researchers to extract more from papers they've read by providing quick search for all their documents with AI-enhanced Q/A and summarization.

In an era where information overload is the norm, Second Brain emerges as a pioneering platform, transforming the way individuals manage, access, and leverage their knowledge assets. Breaking down our unique value proposition a little further, we can talk about the core advantages and enhancements we provide for the research experience.

1) First, we give summaries that make it much easier to understand where key facts are for each paper you've read. Second Brain enhances user interaction by providing concise summaries of documents, facilitating a quick grasp of essential insights without the need to dive into the full text immediately.

2) Our second advantage is the centralized and hyper-usable storage space for your documents. By allowing both uploads through pdf to the website, and chrome extension HTML pass-through, we make sure everything you've read is in one place.
3) We further add smart-clustering to build out folders so you don't have to worry about categorizing the documents if you're struggling to do so yourself. This clustering even allows for you to find new connections between papers that you may not have originally caught.
4) Next, we provide fast search capabilities through the entire contents of every document you've read, which enables you to find even the smallest references you need to get to.
5) Finally, our chat features such as Q/A allow you to answer questions that you otherwise wouldn't be able to answer. Consider, for example, asking a question such as "which human ancestor has the first mutated sequence of the MYH-16 gene?" Google directs you to a few papers, of which, only the third will tell you that the australopithecus was the first ancestor with the noted mutation in their genome. If you have the papers downloaded, you have to take the extra step of remembering which paper would possibly mention this seemingly small detail. However, rather than spend 15 minutes parsing different documents as I did, with the original documents already in Second Brain, you could simply ask our chat and it would answer in seconds.

Due to the strength of our value proposition lying in research, our go to market plan would involve targeting universities and graduate students with our freemium model. We'd aim to draw them in with the simple chrome extension and have them use our website from the time they start their research. This would encourage them to spread usage to other people and even stay on our platform as they end up in industry roles, where we again hope to grow. Our commitment is to not only facilitate information storage but to revolutionize the intellectual workflow, making Second Brain an indispensable ally in the pursuit of knowledge and discovery.

Customer Segmentation:

As our company grows, we hope to target researchers, both in industry, and in academia, who parse through thousands of hours of content. Our ambition goes above the conventional boundaries of information organization and access -- we aim to empower a diverse spectrum of researchers to harness their full potential.

Currently, we are targeting a bifurcated customer segment. First, we're going to target researchers in STEM fields with an early-adopter lifestyle. These will likely be younger, first or second year, researchers (either Masters or PhD students) as they'll benefit the most from easily searchable research. From navigating the complexities of new research domains to digesting an ever-growing corpus of literature, these individuals face a steep learning curve. These students will struggle from key pain-points such as a lack of in-depth understanding in the field, a need to read and understand a large number of papers every week, and the struggle of working in new fields, essentially making google useless without large amounts of time invested. Our unique

value proposition aligns perfectly with these pain points, as we make it easier to get context in any field, summarize documents for in depth understanding and recall, and we provide Q/A that lets the user search things in their knowledge base quickly. Second Brain emerges as the leading solution that aligns seamlessly with their academic endeavors.

In order to expand and develop to the point where we're going to work with these students, we'll narrow our target further to a beach-head market of masters and first year PHD students studying STEM fields in Philadelphia. We find that this group is the most easily accessible for us, and with a research hub for medicine in our city, it will be easy to find any number of labs that would need to enable their researchers. As we expand out of this beachhead market, we'll branch out to other research hubs nearby, and really target advertisements with Google scholar ads. We'll also look to have professors use and endorse our product so they can have their students use it and provide feedback.

Our second target customer is researchers at biotech companies. These R&D focused companies such as Merck, Eli Lilly, and Pfizer, employ thousands of people who are responsible for countless documents and a myriad of information. We aim to be the leading workflow tool for these employees to make them at least 40% more efficient and to help them deliver the highest quality of work possible. In an industry where the pace of innovation and the accuracy of data is non-negotiable, Second Brain offers a solution. By streamlining access to information, enhancing the organization of documents, and facilitating swift, insightful analysis, Second Brain is a catalyst for company success.

Market Research and Sizing:

Our initial market research was sent in order to tailor to our target market more and get a real understanding of what they were looking for. As we came into the project, we assumed that lawyers would be interested in our product, however, we quickly saw that there was much more interest from researchers in STEM fields, such as medicine and Computer Science. They read more papers per week on average, and were generally much more excited about the summaries, with over 85% of respondents saying they would be interested in auto-summarization as a feature.

As we've built out the product, we've had user interviews to discuss how our customers would like to view their research and work through their folder system. From these interviews, we decided to iterate on our original figma design and have finalized an overall frontend framework with user feedback incorporated. We've been tweaking the design consistently to enhance the user experience and will continue to consider UI/UX a priority for our product.

We've also taken the time to start preliminary market sizing from a bottom up approach. From our interviews, we've concluded that our average user will be willing to pay about $70 per year, as they would for a Spotify premium account (or bring in that amount through targeted ads, which will be discussed in the revenue model). Given that about 200,000 students get a PhD every year, with 6 years worth of classes, hoping to have all PhD students using our product, we

are looking at a market size of 1,200,000 people. With $70 per year, our total acquirable market has a $84 million valuation. When we narrow this down to the first couple years of PhD's, we see a valuation of around $33 million. Segmenting this further to our target market of STEM PhD's (which constitute ⅔ of all PhD's), we find a strong initial target market size of $22 million, with lots of room to grow.

Next, the number of employees in research roles proves to be very promising for Second Brain. According to CBRE, there were around 550,000 industry based research roles in the United States in 2022. These positions increased in frequency with a 3% CAGR over the years 2002-2022, implying steady growth. With a boom in AI implementation and research, Second Brain is predicting these roles will only get more prominent. Hence, we feel confident in our ability to expand out and capture this ever growing market with a size of $38.5 million per year.

Stakeholders:

Second Brain's main stakeholders consist of our customers and Cloud Providers and Third Party Services.

Customers:

As discussed above, our main customers are researchers and research hubs/labs/companies. On the individual researcher side, they will be able to interact directly with the software and will save a significant amount of time by not having to manually catalog, document, and summarize all of the articles they have read manually. Instead, using Second Brain, they have all of the information stored for them and neatly organized on our application.

On a larger scale, research organizations stand to gain immensely from our licensing model. By integrating Second Brain, these entities can achieve unprecedented savings in research and development time, catalyzing both innovation and profitability. These organizations will be the one to pay for the service, and we will deliver them an exponential saving in R&D time, and therefore help grow their profits. In this case, we also have the end users (researchers) as our stakeholders as well.

Thus, our solution presents a dual value proposition, serving both the direct users—researchers—and the organizations that facilitate their work.

Cloud Providers and Third Party Services:

Second Brain relies on several different suppliers who act as stakeholders in the platform as well, specifically for hosting our platform and for the custom LLM. Through strategic partnerships with AWS, Netlify, and GitHub, we ensure that Second Brain is not only scalable and secure but also continuously evolving to meet the demands of a dynamic research landscape. On the LLM side, we are using Vespa for our hybrid search engine and using Mistral 7b and Llama2 for the model. We also heavily integrated with the Google Developer APIs to develop our Chrome Extension, and streamline our Google Oauth login flow.

Lastly, our other stakeholders are: 1) our group, who have been actively involved with developing our product and 2) the media, for whom we will need to advertise our product and outreach to sell ads on our application to. Together, these stakeholders form the backbone of Second Brain, each contributing to a shared vision of making research more accessible, efficient, and impactful.

Competition:

As with any company, our unique value proposition separates us from our competitors, but we still have competitors with key strengths and weaknesses.

Our primary competitor is Rewind.ai. This company records your full computer screen, and it uses this recording to synthesize all the documents and information you access over time. It stores this information in a knowledge graph that it uses to allow you to search similarly to us and it provides similar Q/A work. While they are strong in their privacy and security, running completely on the computer with the app downloaded, they struggle in general application. They require a third of a core be dedicated to their application at all times for background tasks, and many of the smart clustering and search capabilities we enable aren't provided as of yet. Furthermore, they're still in the early stages of development, and we are unsure of their market fit. Recording every moment of your screen time seems excessive in most cases, and we are curious to see how users react to the application.

Now, we can't ignore the fact that this product faces competition from home-built solutions. Many users currently store all the articles they read on their hard-drive, and some upload to cloud storage such as google drive. Others will use PDF parsing software to search through everything they've downloaded. While this makes it easy for them to find articles, it requires that they have a strong folder system in order to search their documents, or they otherwise rely on their respective search tools. There is no automatic folder creation/clustering capability and no Q/A. Also, you can't easily download web-pages. Our goal is to address the pain-points associated with these current methods of document storage in order to enhance the research experience for anyone who needs to understand their field better.

Revenue Model:

After much discussion in the group, we've decided that the best revenue model is likely going to be a 'freemium' model, similar to Spotify, with a licensing option for enterprise. For the freemium model, the logic here is that a large number of our initial user-base is going to have limited income to spend on subscriptions. However, they're going to have a lot of data on interests in their databases, so we can send targeted ads their way. We can charge the user a subscription price that roughly correlates to slightly more than their revenue to us from targeted advertisements should they wish to get away from the advertisements, and we'll guarantee profit. Additionally, for users utilizing our free model, they will be subjected to a rate limit on the

number of queries they can ask our model. For users in our paid model, their cost per month will scale as they accumulate more articles and contribute more to our costs for storage.

Based on the current subscription prices university students are willing to pay for applications such as Spotify, Netflix, and Hulu, we're aiming for a $70 per year subscription. The logic here is that these are subscriptions that students pay a discounted fee for and they tend to have no real necessity in anyone's life, so they should be comparable in willingness to pay from the sheer convenience they bring in. In addition, we would offer different models of pricing where you pay less per month for the longer your investment.

As for mapping this subscription cost to our cloud charges, we also noted that the amount of data stored for an application such as spotify with countless songs on countless servers in addition to information on which playlists each song is present in will roughly equate the number of articles our users maintain in their Second Brains, so understanding that spotify profits from the students significantly, we know that having a base subscription cost in a similar range should work well for us. We also plan on adding costs for cloud compute that people bring in if they exceed a base value. This would ensure we cover our costs and even potentially earn a cut on the excess compute.

Add in the fact that our data on people's research interests will be vast and accurate, allowing for precision targeted advertisements which we can charge significant amounts for, and I think we'll have a safe series of revenue streams.

Cost Model:

The majority of our costs will come from cloud compute and storage. We're planning on using AWS to host our models and store the majority of our user data as we start our company. We understand that this dependency will lead to scaling costs as our business grows, but we also plan on using our subscription model to allow for rate changes that ensure profit for us as we scale. Further, as we grow, we could shift to our own servers and host our models for cheaper after the initial fixed cost investment. Other than that, our core costs are simply personnel, making this a very low cost endeavor. We won't need factories, manufacturing material, and we may not even need office space.