Summer Midterm 2023
Clairvoyant

Congratulations! Based on your enrollment in a MSBA program, you secured an internship with Capital LLP, a prestigious early-stage venture capital firm specializing in SaaS application investments. During your internship, Bob Partner, a seasoned professional, is your mentor.

Your first assignment involved meeting with CodeTech founders, Betty and Pierre, who possess advanced degrees in data science. Bob advised you to take thorough notes during the meeting, and you diligently followed his guidance.

After the meeting, Bob expressed his interest in the opportunity but also raised nagging concerns about the potential ethical, security, privacy, and risk issues associated with CodeTech's inaugural offering, Clairvoyant. He asked you to write a memo explaining the key ethical, security, privacy, and risk issues that you have identified with the opportunity. Bob needs the memo by Monday at 6 p.m., as the firm's investment committee convenes on Tuesday morning.

You have heard from fellow interns that Bob requires a concise and focused memo. Bob expects a maximum of five typed pages in 12-point font due to his visual impairment. In addition, he appreciates a well-structured memo that presents a thorough discussion of the facts.

Your notes from the meeting with Betty and Pierre:

Background

CodeTech is a nascent software development company founded by Betty and Pierre and located in San Francisco. Their mission is to develop tools that empower individuals in the realm of developing friendships. Their first product, Clairvoyant, is currently in beta testing and has received initial funding from family and friends. The founders describe Clairvoyant as a friendship augmenting SaaS based application, that when fully developed can be used widely across a broad spectrum to identify individual traits: from whether someone is generally happy, sad or self-confident, to their political perspective, or even their sexual orientation.

Based on research following the pandemic that indicated that many individuals identified as being lonely and isolated, Betty and Pierre believe that there is a market for non-traditional friendship applications (as opposed to dating applications), that enhance a user's ability to meet people on broad based community or professional platforms (e.g., Nextdoor, LinkedIn, etc.) that share a trait they desire in a friend.

After reading several University research papers on facial recognition (including research that neural networks can infer/detect sexual orientation from facial images), Betty and Pierre realized that facial recognition models can be used for more than security or identification. From pictures, models are increasingly able to identify individual traits. However, they also understand that the accuracy of facial recognition models depends on many factors, including the quality of pictures, which is outside of their control. Their epiphany came when they realized they could supplement a facial recognition model with a natural language processing model to assess not just a picture, but also the text (e.g., a profile, comments, etc.) that accompanied most pictures associated with community or professional platforms.

The First Target Audience: LGBTQ+

Both Betty and Pierre took a business course from a local high tech business incubator nonprofit. The course taught them the importance of the principle that a business should identify an initial target audience before attempting to expand its market. Focus is imperative.

Betty and Pierre decided that they should initially target LGBTQ+ individuals as consumers of their application. Their research found that many LGBTQ+ users of existing applications (e.g., Discord chatrooms, Reddit community boards or applications like Bumble) who are seeking friendship connections were dissatisfied. Betty and Pierre further noted that research by Meetup.com found that the number one search term in 2023 for meetings on its application was "Friends" and that the number five search term was "LGBTQ" (compared to 135th in 2022). See Meetup Measurement Report, 2023, p. 8. From the study (and other research), they surmised that many LGBTQ+ individuals were looking for friends who shared a common sexual orientation.

They also found that many users of sites like LinkedIn and Nextdoor do not indicate their sexual orientation in their profiles, and that while it might be possible for an individual (without an

application) to infer sexual orientation by other means (e.g., searching on the organizations an individual belongs to), it takes time, and it is inconvenient.

The Demo of Clairvoyant

Clairvoyant is a proprietary deep learning algorithm, that includes scraping software that can be targeted at a particular site. From the demo of Clairvoyant it is clear that the user (also referred to as the "Subscriber") interface is simple: (a) the user simply clicks on a box of a commonly used sexual orientation name, (b) picks a website from a list, and (c) clicks on those predefined parameters that the user wishes to include in their search (e.g., zip code, profession, etc.). In our demo, we asked Clairvoyant to identify all individuals having a particular sexual orientation on a site within a particular zip code and got a response within seconds.

Clairvoyant only supplies to the user the identity of those individuals who its proprietary model determines have an 80% or higher probability of having the sexual orientation requested. After receiving the results of the search (and not through the Clairvoyant application), users can then reach out to the individuals that were identified through the search, initiating conversations without the formalities or pressures typically associated with traditional meeting/dating applications. If a Clairvoyant user decides to reach out to an individual identified through the search results, it is up to the Clairvoyant user to decide how much, if any, information regarding their sexual orientation (or their use of Clairvoyant) they wish to disclose.

To overcome sites that have username and password protections, Clairvoyant requires that its users have an account on the site that they want to search (e.g., Nextdoor, etc.) and provide their account and password information to Clairvoyant for that site. Clairvoyant uses that information to enter the protected site to scrap the data and run the requested search.

Clairvoyant does impose limitations on the scale of the search for several reasons, including it is possible to harm a site (e.g., LinkedIn, Nextdoor, etc.) that it searches if the search is too broad and/or vast (e.g., the search might degrade a site's performance which results in a type of denial of services attack). To date, since Code Tech has not had the resources to sufficiently test the application, it has imposed the search limitations at a level that they feel "comfortable" with, and "are pretty sure well not cause harm."

Initial Revenue Model

From a revenue stream perspective, CodeTech sees two initial major sources: (a) subscription fees and (b) revenue from selling its data through data brokers. Its plan is to initially charge a minimal monthly subscription fee to gain as wide an audience as possible. From a subscription basis, they plan to first focus on the United States and European markets since Pierre is French.

After they have acquired sufficient subscriptions to have enough data to be attractive, they plan to sell their data through data brokers. The value of LGBT+ information can be significant to data brokers and advertisers, as it represents a specific demographic group with distinct characteristics and interests that can be targeted. However, it is not clear what information data brokers will eventually be interested in, and as a result, CodeTech's initial strategy is to retain as much data as possible.

Presently the data stored by Clairvoyant falls into two broad categories:

1. Subscriber Data. Subscribers are required to provide their name, gender, age, address, and email address, create a Clairvoyant username and password (there are no password restrictions - e.g., length, complexity etc.), and provide their username and passwords for any protected sites they wish to search. Each subscriber's queries (e.g., the search terms), along with the results of the searches are also stored.
2. Search Data. CodeTech retains all search results indefinitely. While only those results that are scored ≥ 80% are provided to a Subscriber, the results of searches stored include all individuals searched within the search criteria, including their scores, along with other sexual orientation corroborating information found on the site (e.g., sexual orientation self-disclosers, organizations, etc.). Thus, even if a search result only indicates that an individual has a 10% probability of having a particular sexual orientation, that information is retained by Clairvoyant (although it is not shared with the Subscriber).

All of the data is stored on servers in the United States.

Privacy and Security

Because of funding limitations, CodeTech has not prioritized privacy. Instead, it sought the advice of friends who operate Utah Data, a tech company that provides services solely in Utah. With the blessing of Utah Data, Betty and Pierre borrowed Utah Data's privacy policy, which is now prominently displayed on the Clairvoyant site. I got a copy of the policy from Pierre (see below).

At the moment, CodeTech is not overly concerned with security since it uses ARCP, a reputable third-party cloud provider. The responsibility for information security is outlined in a Security Agreement between CodeTech and ARCP, which states that ARCP provides physical, infrastructure, platform and network security, along with certain backup and disaster recovery mechanisms. ARCP has also offered CodeTech options for data security (e.g., for a price it can encrypt data), which CodeTech has declined based on cost. For those aspects of security that ARCP is responsible for, it has provided CodeTech with independent audits and certifications to demonstrate compliance with industry-recognized security standards and regulations.

Application and data security are expressly delineated in the Security Agreement as the responsibility of CodeTech (e.g., user authentication, authorization and access, etc.), and not ARCP. ARCP has suggested that the security of Clairvoyant be assessed by an independent security third-party. CodeTech did have a preliminary security assessment done of Clairvoyant by SSF, a small security firm, to see what level of effort and cost it would take to get an independent security certification based on common security standards and regulations. Betty and Pierre thought that the security deficiencies identified by SSF would cost too much to remediate. Betty said, "SSF treated CodeTech like a giant entity. They expected us to have a risk assessment, written security policies and procedures, and all sorts of controls like intrusion detection and monitoring systems. We don't have time for that bureaucratic nonsense. They even expected our developers to be trained in secure coding. Give me a break – we're a start-up!"

One bone of contention between Betty and Pierre is whether they should submit their Clairvoyant model to outside testing for accuracy. Pierre is vehemently opposed to the idea and to date has vetoed it. Pierre is extremely concerned that the intellectual property value of their proprietary model must be protected at all costs. To quote Pierre: “I created the model. I am best positioned to test it!”

CodeTech Privacy Policy – Our Assurances to You

CodeTech is the provider of Clairvoyant (collectively, "We"), a SaaS based service provider that assists you (the "Subscriber") meet new friends who share common traits and orientations.

- We collect Subscriber personal information only to provide our services, improve Subscriber experience, and or fulfill our legal obligations.
- We do not share personal information with third parties that are not associated with providing our services to our Subscribers.
- Subscribers can be assured that we will never sell your personal information.
- Subscriber information is only retained for as long as it is necessary to provide services to you, improve user experience or fulfill legal obligations, after which we take reasonable steps to destroy the data so that it cannot be accessed or misused by any unauthorized party.
- We take all reasonable steps to ensure that any personal information we collect, or share is accurate and suitable for the services that we provide to you.
- We take all reasonable precautions to protect personal information and we regularly monitor and test our systems and procedures to ensure our precautions work.

If you have any questions, please contact us at wevalueyourprivacy@codetech.com