Big Duo is Watching You

An exploration of Duolingo's Privacy Policy and the dangers of obscuration

Lyd Wyble, for WRIT 312

Like any social media platform, we know that Duolingo is constantly collecting data on our activity. But how exactly this is done—and with whom the data is shared—is a **mystery** to most users.

As with many online platforms, the "issue" with Duolingo's Privacy Policy lies **not in omission, but obscuration**. The policy is existent and well-explained, but it is **difficult for users to find and interact with.**

What's actually collected?

When you create an account:

- Name and Age
- Email Address
- Phone Number
- If you log in with Google or Facebook, Duolingo receives info from that service (such as contacts and ad preferences)

Each time you log in:

- Device/ browser data
- IP address
- "Data regarding your engagement" (time spent on app, probably)

Displayed on your profile:*

- Your stats (total exp, league, etc.)
- Streak (indicates day of last app usage)
- Friends you have, and users that have added you as a friend

***All of this information is "Public" by default,** but can be hidden by setting your profile to Private.

◀)) The price of publicity:

While public profiles allow users to explore the full potential of Duolingo as a social media, they also allow **web scrapers** and other third-party sites to "use your public information for their own purposes." That sentence doesn't take any accountability, so it is a fair assumption that our **data is then no longer protected by Duolingo's Policy**. One example of this would be user posts and comments being used to train LLM generative AI such as **ChatGPT**.

4

Speaking of ChatGPT…

If a user subscribes to paid service Duolingo Max, they can send text and audio messages to an AI conversation partner. This information "may be shared with AI vendors such as OpenAI, which may keep a copy of your text or audio responses for their own purposes." Users are not re-presented with the Privacy Policy upon subscribing (or presented in the first place… but more on that later).

Aside from Duolingo Max, user-recorded audio from speaking exercises is sent to third-parties like **Google** and **Amazon Web Services.** This is reportedly for speech recognition, but these are also top names in the world of personalized ads (i.e, these companies really really want your data). Duolingo also collects speech records for "**product improvement**"-- users can opt-out of this, but not the third-party sharing! Pretty weird, right?

Other Interactions:

- [Cookies](): Includes third-party targeting cookies, which track users across multiple websites. Opt-outs can only be done on a separate site.

- Google Analytics: This is another tool that tracks users across the Internet and shares their data. Can only be opted out of by installing a Google browser add-on (weird!).

The FullStory Full Story:

> "A service called FullStory... captures and analyzes your activity and provides us with data and a <u>video session replay</u> of your activity" (underline added)

Unlike a simple cookie, FullStory's session recordings are a method of data collection many users would find off-putting. The service does not record names, profile pictures, or account info, but virtually everything else is logged. The service is admittedly easy to disable from FullStory's website, but I have used the app since 2018 and was unaware of the service until recently.

FullStory is automatically disabled for users under the age of 13. (In a study of 8-12 year olds with social media accounts, 62% had claimed to be 16+.)

Diving In: The New User Experience

In the following slides, I document the experience of a first-time Duolingo user. I wanted to see at what point users are introduced to the Privacy Policy, and to what extent. No real personal information was used in account creation.

Write this in English

Dúo siempre se siente feliz.

Duo always feels happy

more home sorry

I was led through a short survey and first lesson **before an account was even created**. This was likely subjected to the full surveillance of a typical user session, but I was **unable to opt out of data collection and analysis**. (However, the lesson and many questions could simply be skipped).

Create your profile

Age

Providing your age ensures you get the right Duolingo experience. For more details, please visit our **Privacy Policy**.
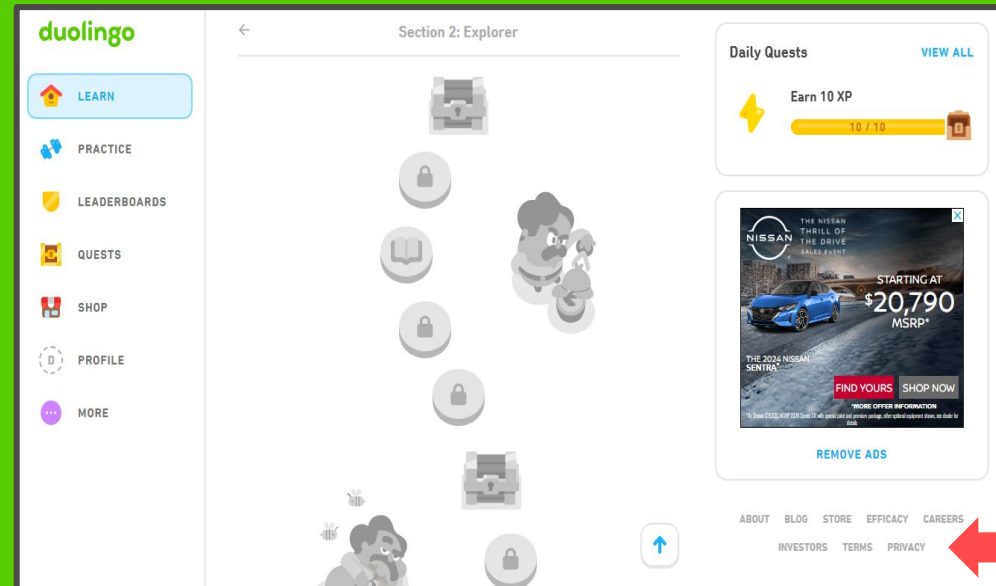
Name (optional)

Email

Password

CREATE ACCOUNT

OR

f FACEBOOK   G GOOGLE

By signing in to Duolingo, you agree to our **Terms** and **Privacy**

When creating an account, we see two mentions of the Privacy Policy (both hyperlinked). Note that this isn't a poorly cropped screenshot– you would actually need to scroll down! Reviewing the Policy is **not mandatory** for account creation.

If a user wishes to view the policy after account creation, it can be found in very small print at the bottom of their dashboard (also requires scrolling)

11

Avenues for Improvement:

Duolingo's Privacy Policy is comprehensive, but it is difficult for users to access and interact with. Some suggestions:

- Present and have users **manually agree** to the Privacy Policy upon account creation; Duolingo could even provide a brief summary and give users the option to read in full.
  - **Re-present the policy** (or summarize relevant changes) when users change subscription plans.
  - **Provide opt-outs within Duolingo's settings,** rather than having users navigate third-party websites.

Users have a right to accessible policies and opt-outs!

🔊 Duolingo's full privacy policy can be found here.

Thanks **for** **listening!**

gracias | | | ありがとう | danke | | Спасибо

Work Cited: "A Third of Children Have False Social Media Age of 18+." Ofcom, 16 Mar. 2023, www.ofcom.org.uk/news-centre/2022/a-third-of-children-have-false-social-media-age-of-18/.

13