Netflix provides a direct contact to the Data Privacy Office via [privacy@netflix.com](mailto:privacy@netflix.com). The information Netflix gathers information users provide, including, but is not limited to, personally identifiable information (PII) such as names, email addresses, payment methods, telephone numbers, zip codes, gender, and age.

Netflix also collects information that is implicitly provided, such as title selections, interactions with our customer service, device IDs (such as a router and other network hardware), and general activity while using the service.

Netflix will also reach out to “partners” such as internet service providers or mobile carriers. The exact nature of the information given varies depending on the partner but can include PII. Excluding the information given by users and partners, Netflix also obtains information from other sources such as advertisers and payment providers.

The purpose of the data gathered is to enhance the service offered, including but not limited to, more accurate personalized viewing recommendations, localized content, increased security, and better customer service.

Netflix utilizes cookies and web beacons (aka pixels) to assist with the information gathering as well as prevent fraud and maintain performance. Netflix also reserves the right to reject requests not required by law. There is a separate section for the California Consumer Privacy Act (CCPA) but it doesn’t vary wildly from the main privacy policy. I do wonder how it would change if my VPN was in Canada or Europe.

Legal regulation is a good segue into the six critical areas of legal risk: data ownership/security, consumer privacy, third-party contracts, regulatory compliance, underlying contracts, and legal discovery. The consumer privacy policy was outlined above.

Netflix has been very clear regarding the lack of 100% guaranteed security. They mention taking reasonable precautions but that is a highly subjective word. On data ownership, there is nothing written down. To me this indicates that Netflix owns the data. It seems even moreso that way when the only mention is that a user can delete their account, but there is nothing written about what happens to the data after deletion.

Third party contacts seem to be prevalent in Netflix’s business model. They work with “partners” and source additional information from other 3rd parties. What each entity provides Netflix is a variety of data in likely a variety of formats. This leads to data loss and potential issues with international regulations – even if the information is going from California to New York, but makes a pitstop at a server in Copenhagen, does that make Netflix susceptible to GPDR rules? Regulatory compliance is rift with stumbling blocks.

Continuing with legal liabilities is the act of discovery. Discovery is the formal process of exchanging information between the parties in a case. With the massive amounts of information being gathered, there are concerns about limiting the scope of the data analysis. This could lead to private, profitable information being shared to a competitor. On a non-corporate scale, personal data can also be lost or leaked to the public.

A concern not listed in the six critical areas of legal risk is the healthcare implementations. Netflix is not a healthcare company but it is predicting the user’s mood by when they log in to the account. Netflix is making predictive models on someone’s state of mind based upon the information they’ve gathered. At some point, it could be fair to say that if the models are accurate then showing a movie about suicide to a suicidal teen would lead to a large culpability on the company’s behalf.

As a data scientist at Netflix much of the information I could use is already being gathered. The harder part is using less private information to get a similar output. To start working on that, I’d try to make everything more efficient. We are getting X columns of data from our various partners, let’s see if we can make it X-1. Variable reduction techniques can lead to a similar output in predictive power with a requirement of less data. This is more secure and less costly to the company. Outside of trying to reduce the dependency on vast quantities of data, I think specific data gathering techniques could be modified to be less private depending on the goal of the researcher. Mood prediction, for example, doesn’t necessarily need to aggregate information from credit card companies along with Netflix specific usage. A hypothesis can be used to classify moods – watching the Notebook at 2am on a Tuesday morning is an indicator that someone has gone through a breakup. We don’t need to know that the user also purchased a tub of ice cream as well. I just realized my paper is 3 pages so I’m going to stop now – sorry!