Software designers and engineers make use of software specifications to design and develop a software system. Software specifications are generally expressed in natural language and are thus subject to its underlying ambiguity. Ambiguity in these specifications could lead to different stakeholders, including the software designers, regulators and users having different interpretations of the behavior and functionality of the system. One example where policy and specification overlap is when the data practices in the privacy policies describe the website’s functionality such as collection of particular types of user data to provide a service. Website companies describe their data practices in their privacy policies and these data practices should not be inconsistent with the website’s specification. Software designers can use these data practices to inform the design of the website, regulators align these data practices with government regulations to check for compliance, and users can use these data practices to better understand what the website does with their information and make informed decisions about using the services provided by the website. In order to summarize their data practices comprehensively and accurately over multiple types of products and under different situations, and to afford flexibility for future practices these website companies resort to using ambiguity in describing their data practices. This ambiguity in data practices thus undermines its utility as an effective way to inform software design choices, or act as a regulatory mechanism, and does not give the users an accurate description of corporate data practices, thus increasing the perceived privacy risk for the user.

In this thesis, we propose a theory of ambiguity to understand, identify, and measure ambiguity in data practices described in the privacy policies of website companies. In addition, we also propose an empirically validated framework to measure the associated privacy risk perceived by users due to ambiguity in natural language. This theory and framework could benefit the software designers by helping them better align the functionality of the website with the company data practices described in privacy policies, and the policy writers by providing them linguistic guidelines to help them write unambiguous policies.