

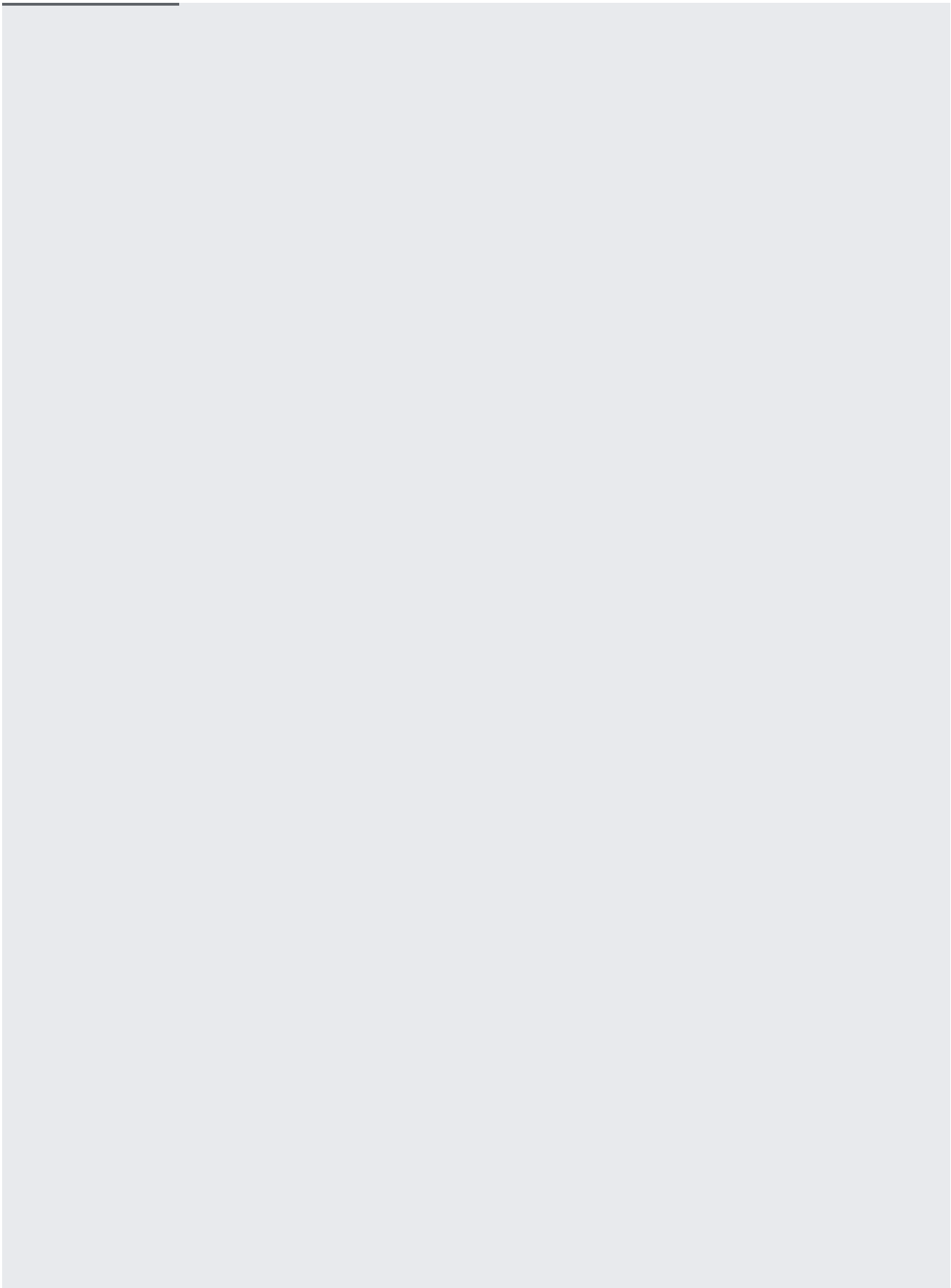
A dataset contains representative samples of the type of content you want to translate, as matching sentence pairs in the source and target languages. The dataset serves as the input for training a model.

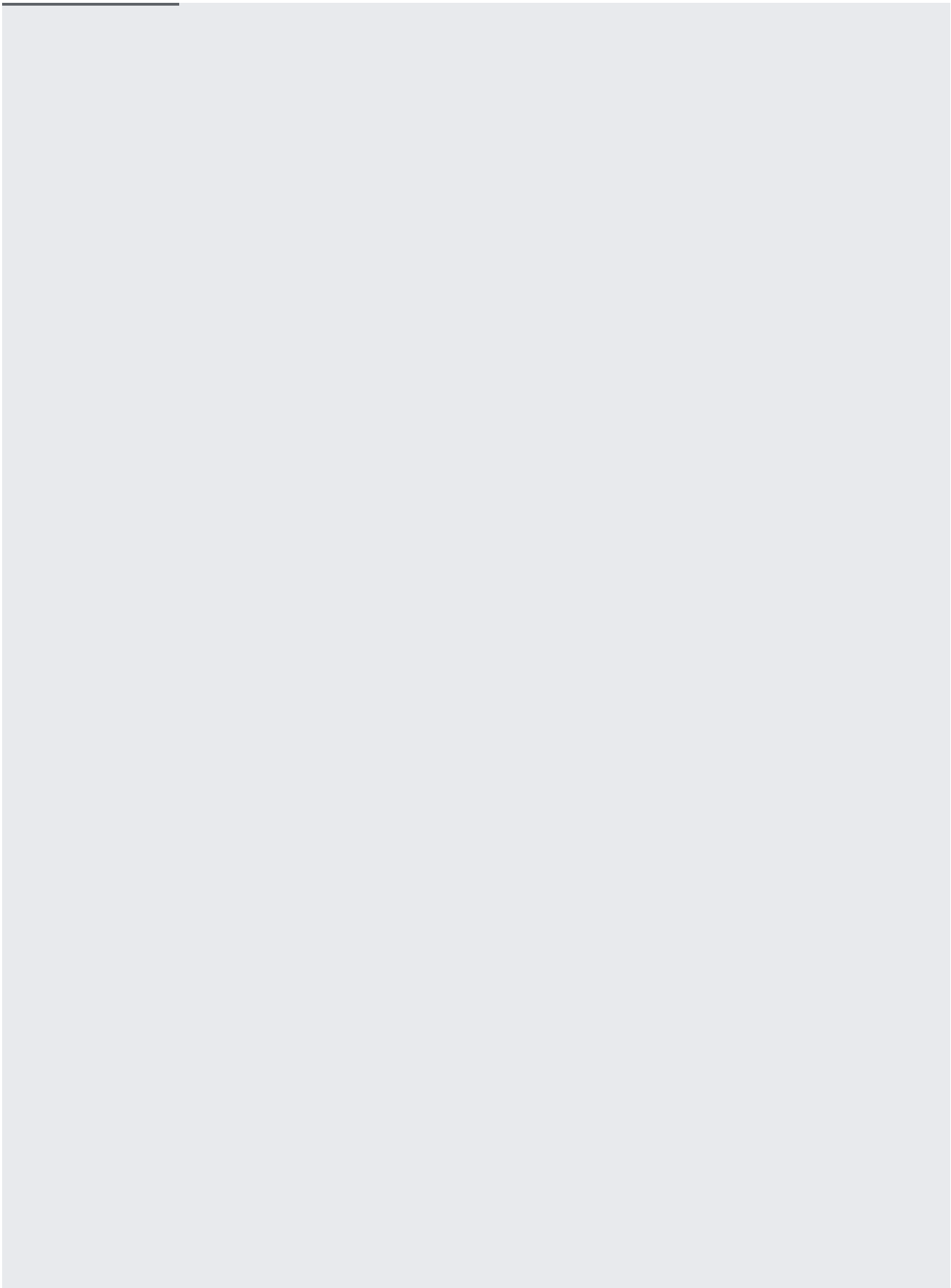
The main steps for building a dataset are:

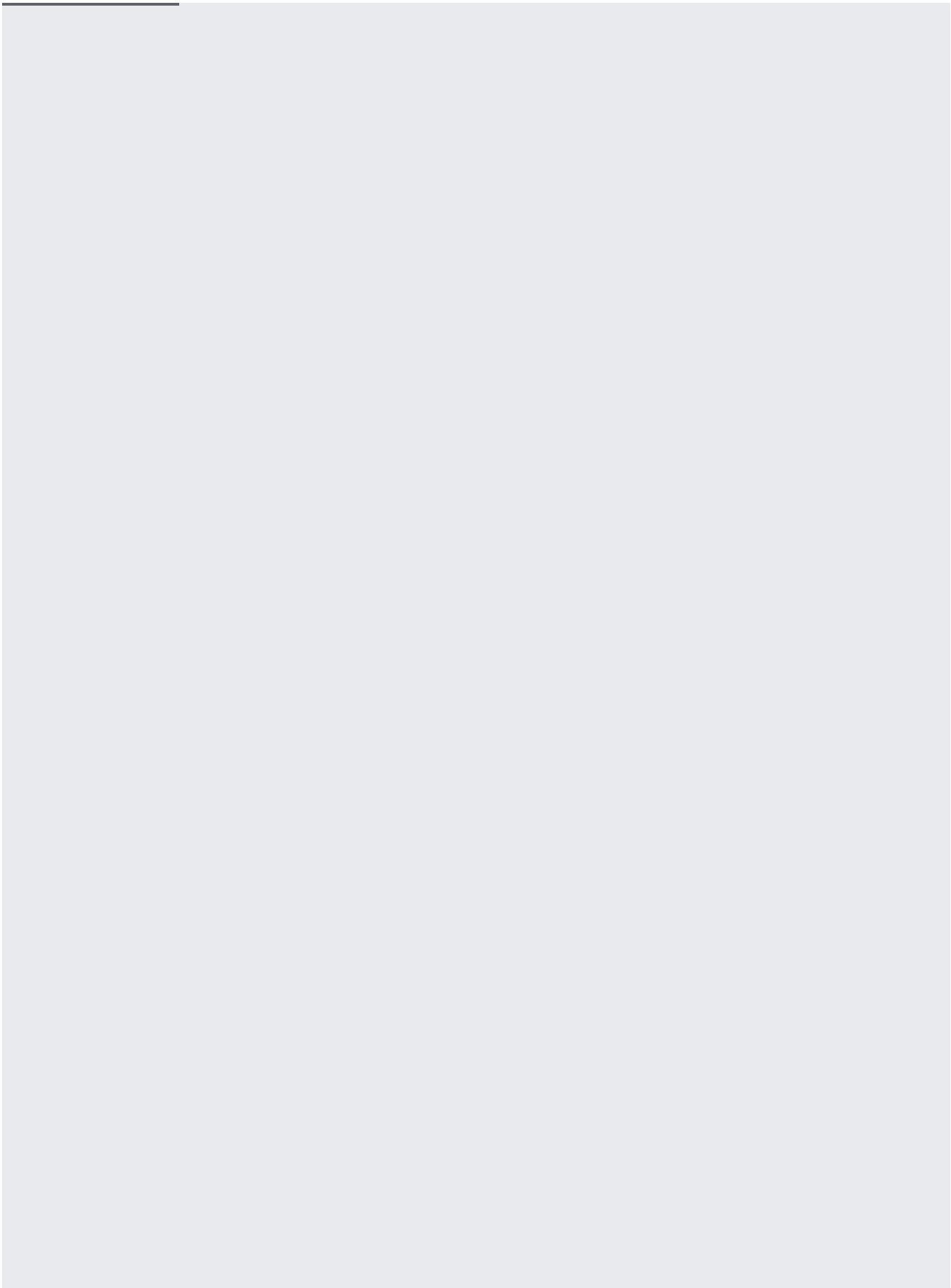
1. Create a dataset (#create-dataset) and identify the source and target languages.
2. Import sentence pairs (#import-data) into the dataset.

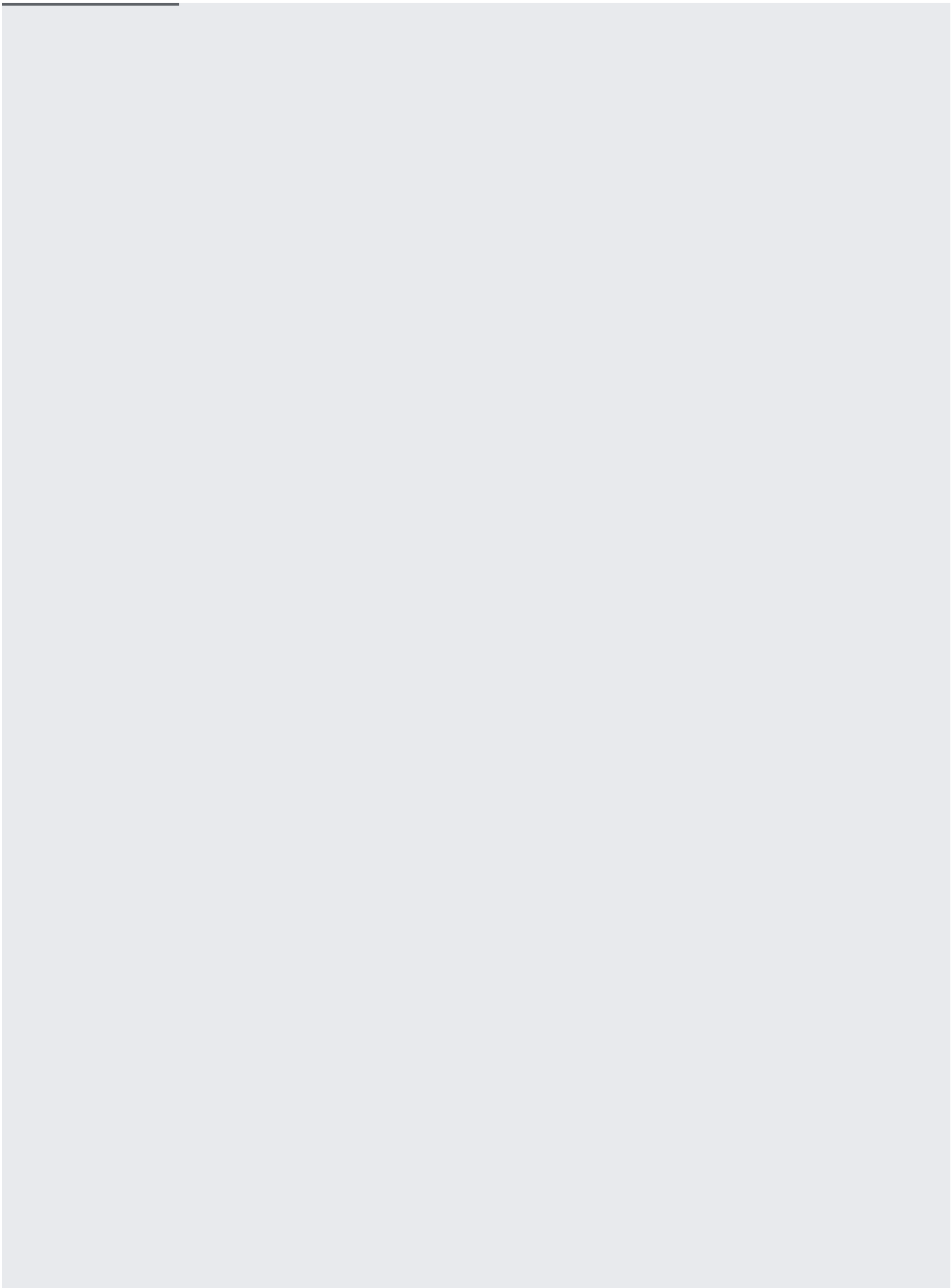
A project can have multiple datasets, each used to train a separate model. You can get a list of the available datasets (#list-datasets) and you can delete datasets (#delete-dataset) no longer needed.

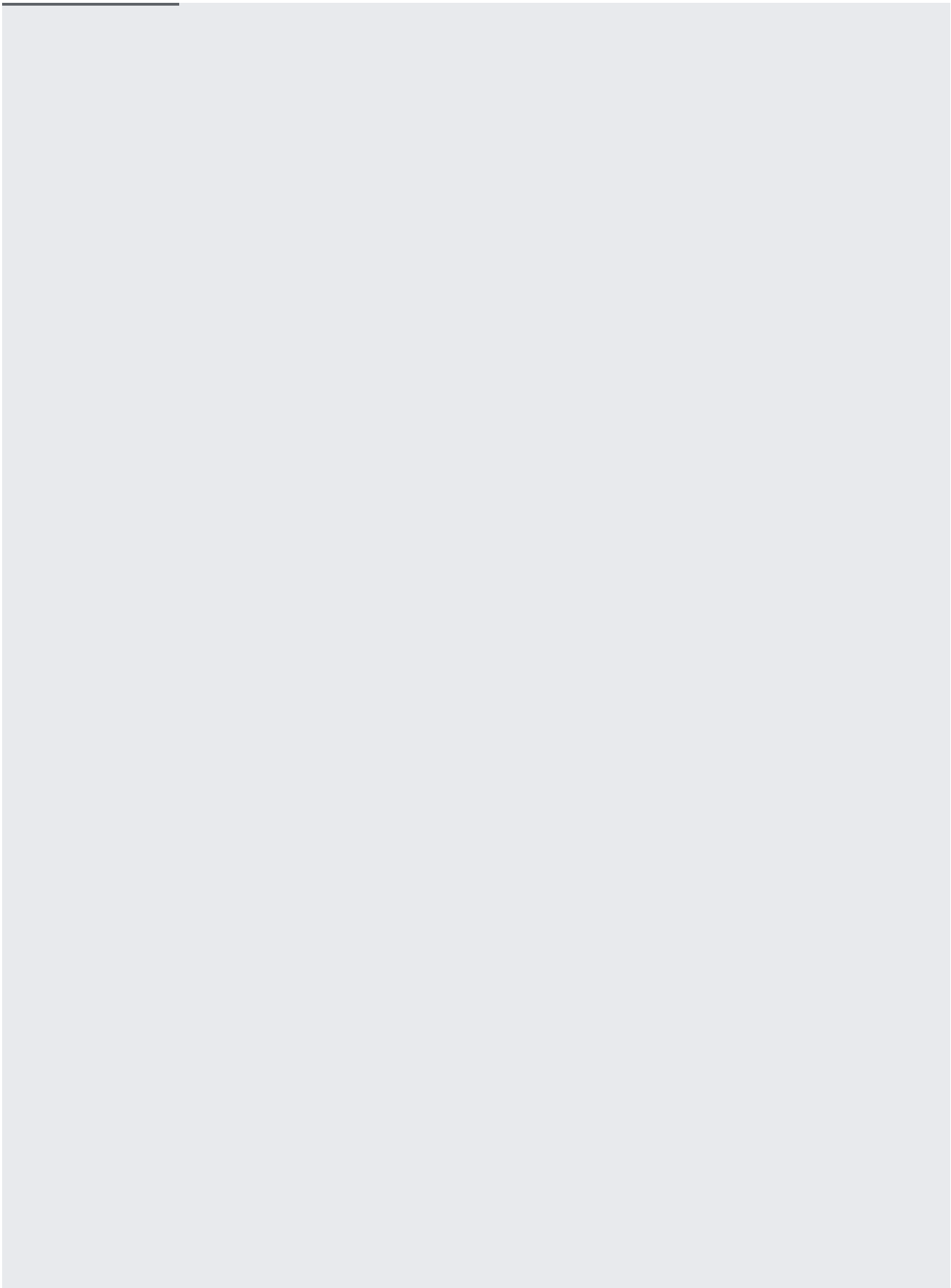
The first step in creating a custom model is to create an empty dataset that will eventually hold the training data for the model. When you create a dataset, you identify the source and target languages for the model.

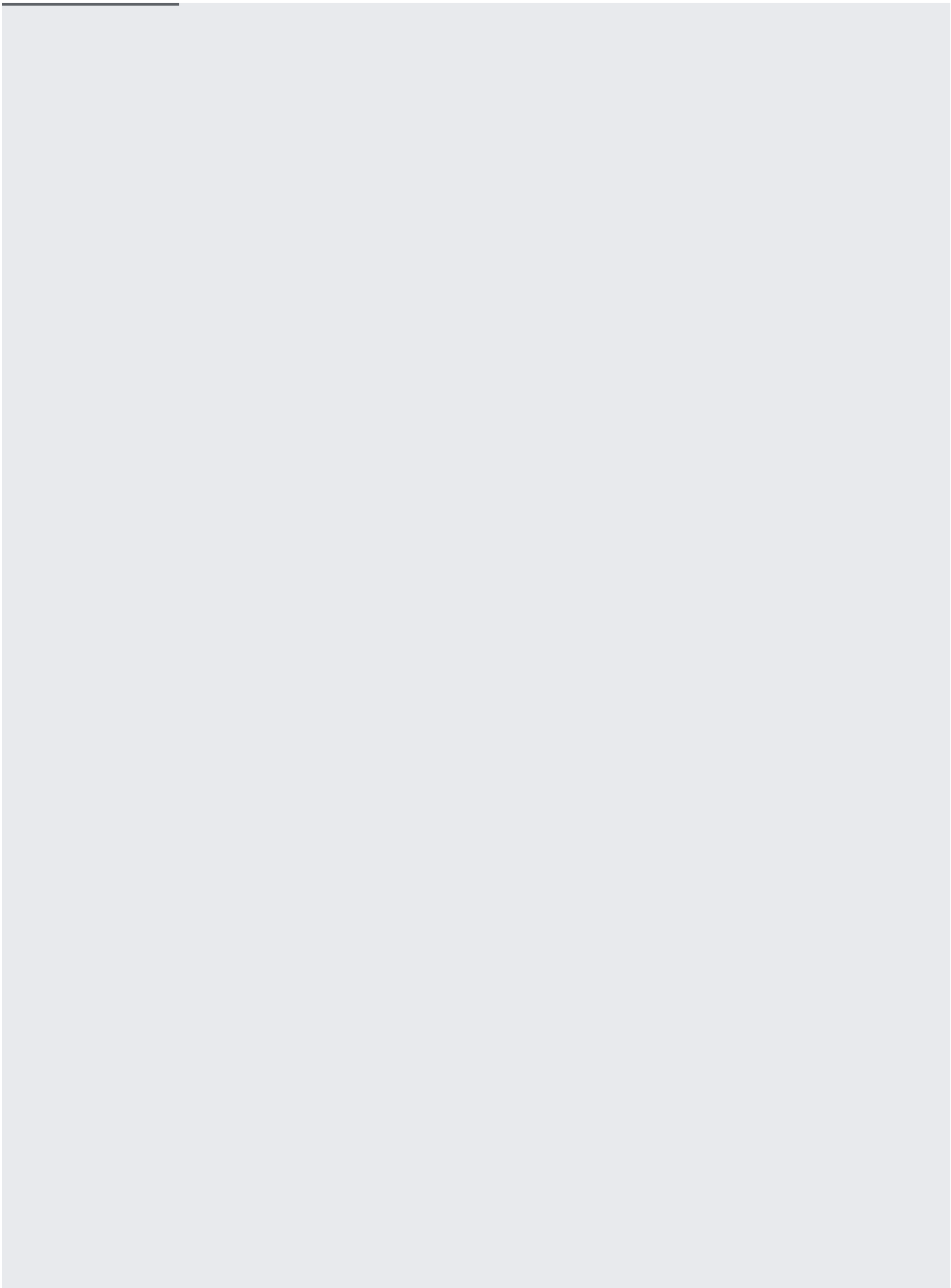


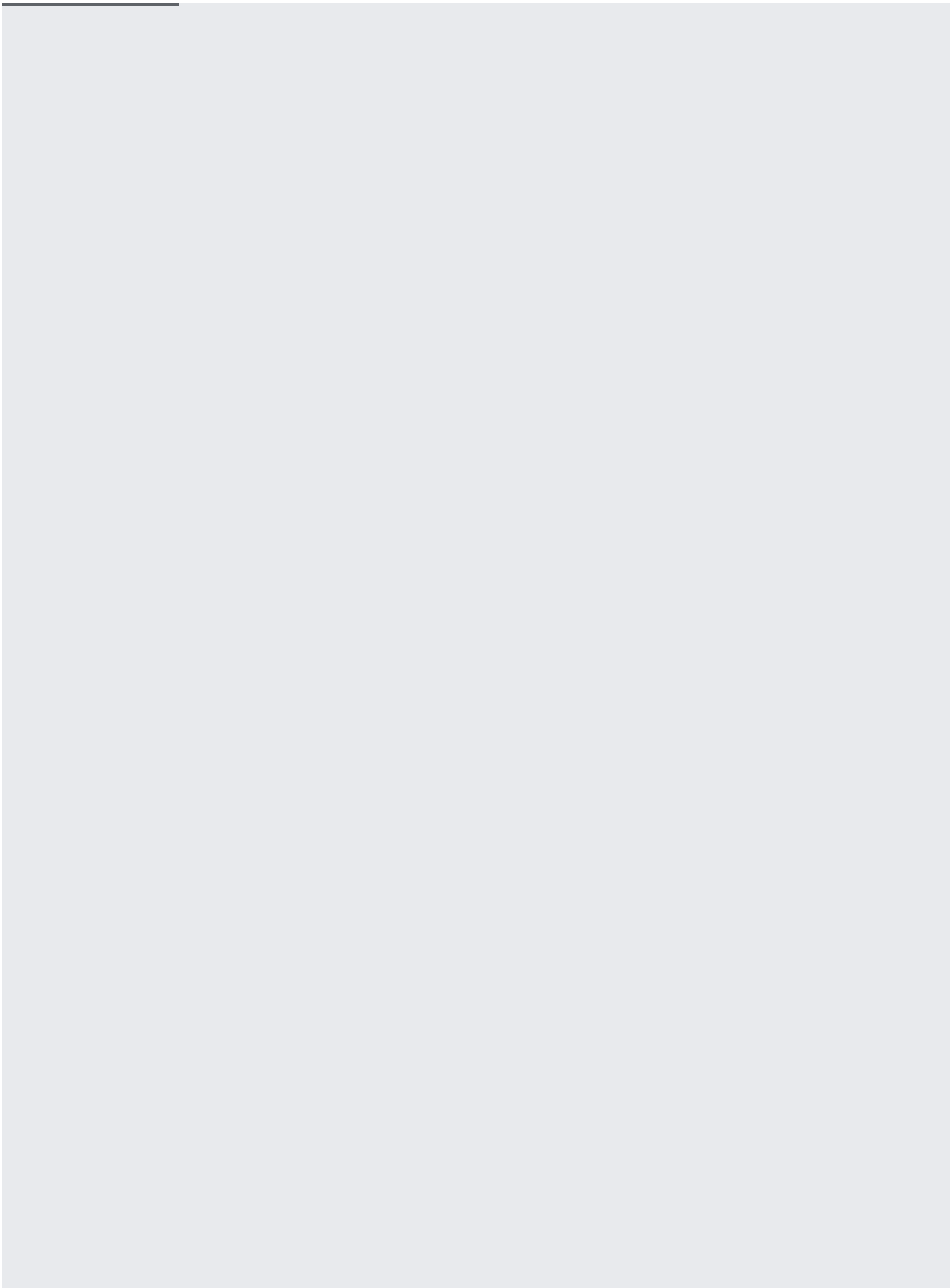




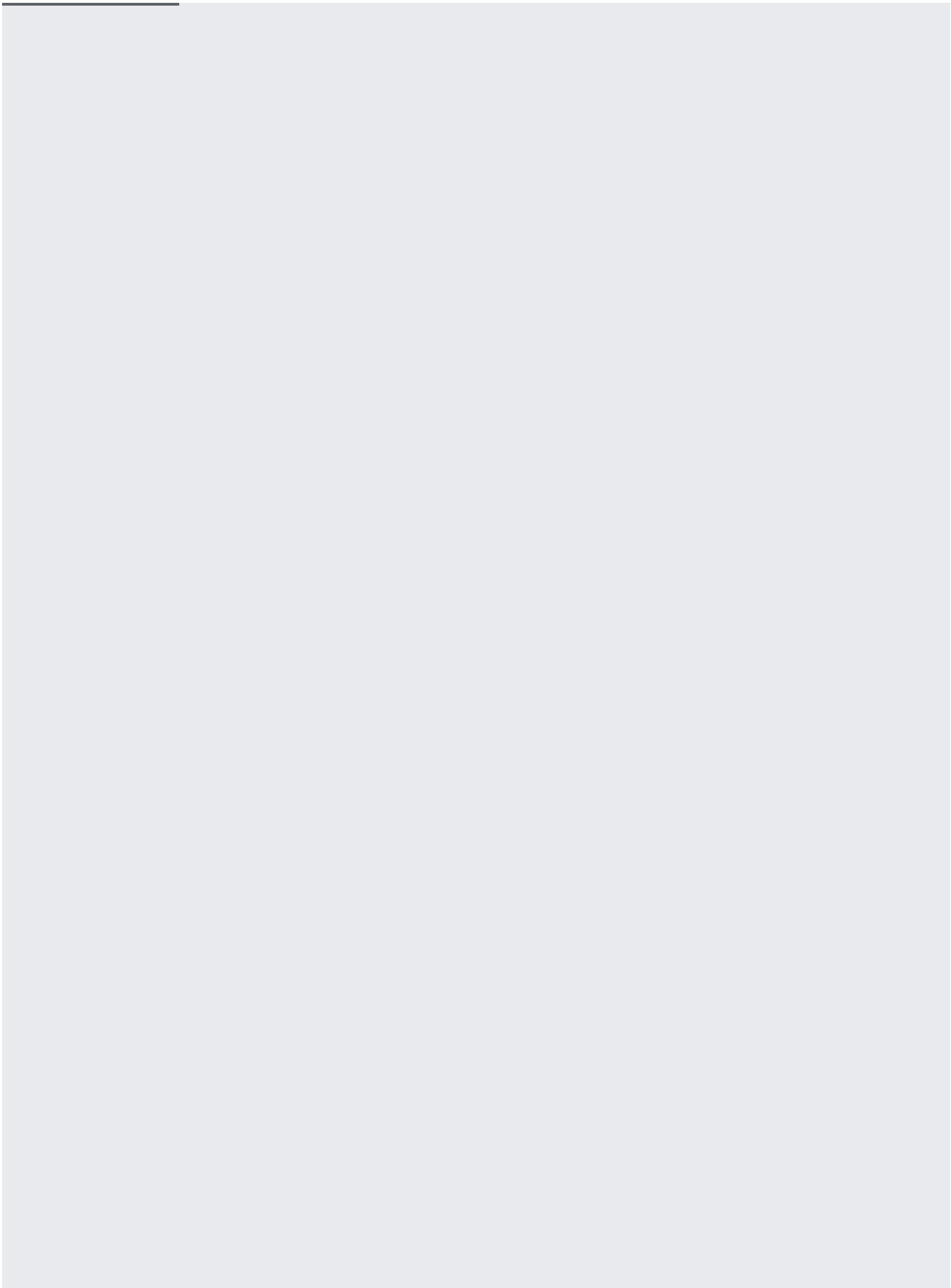


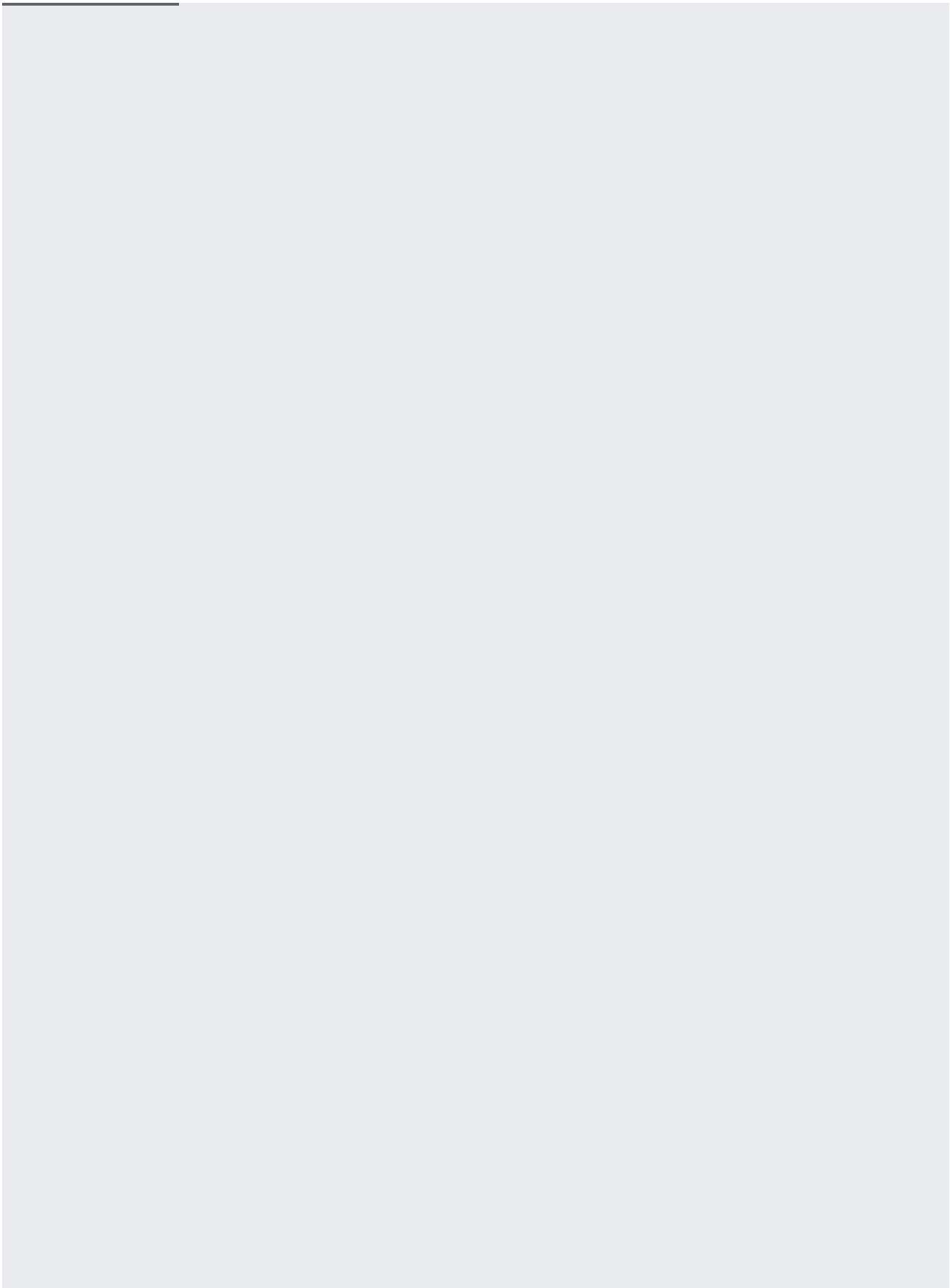


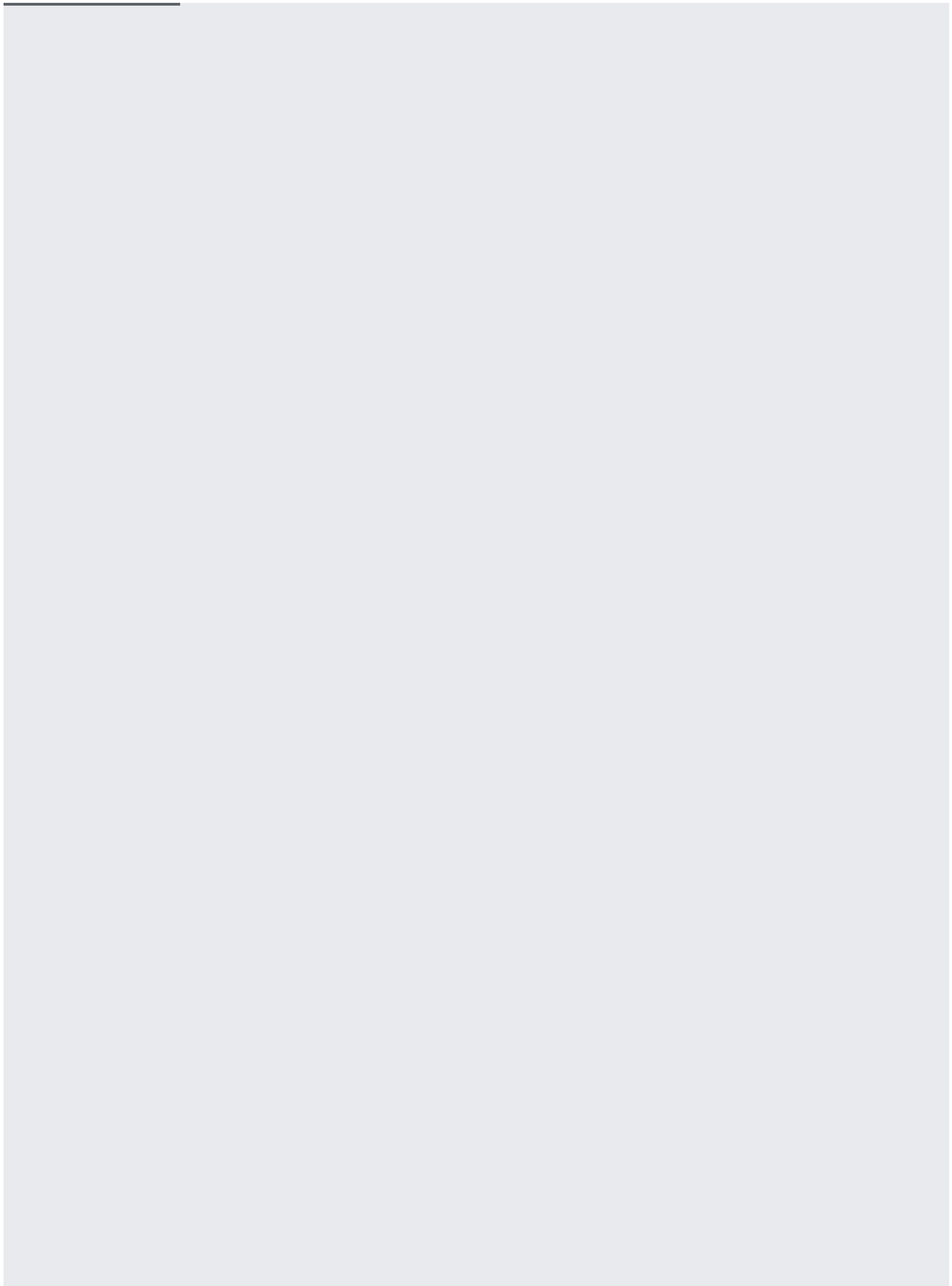


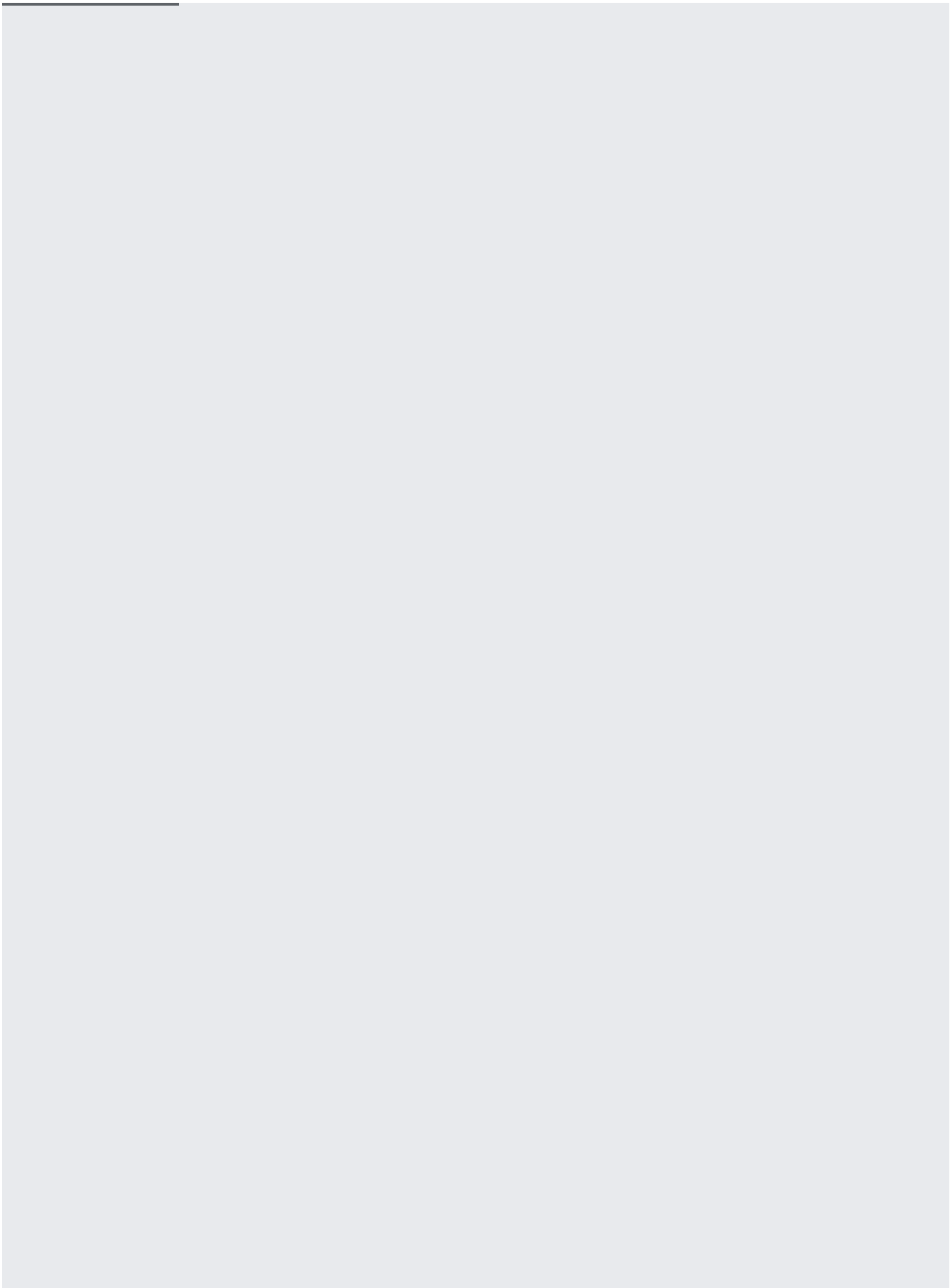




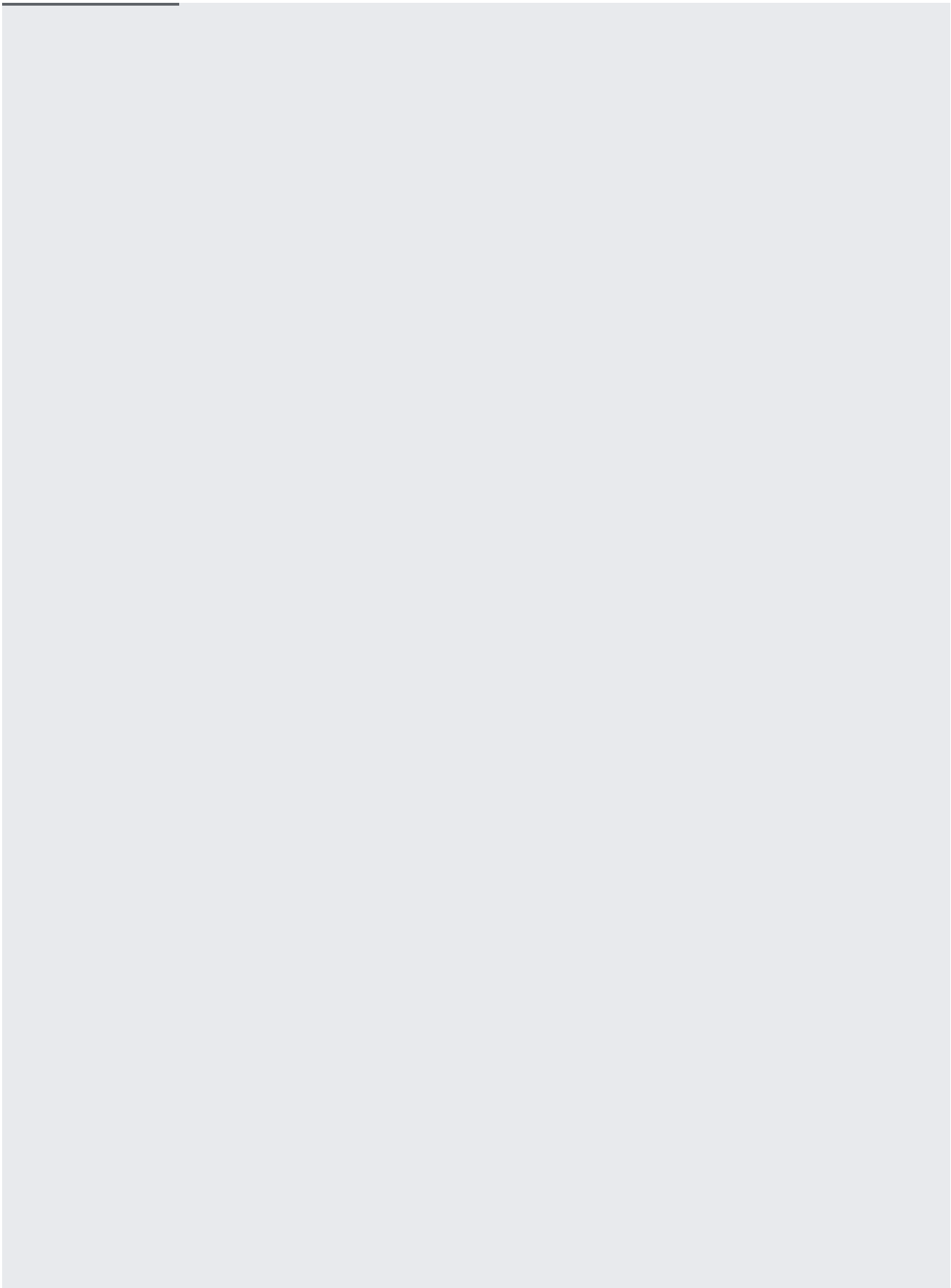


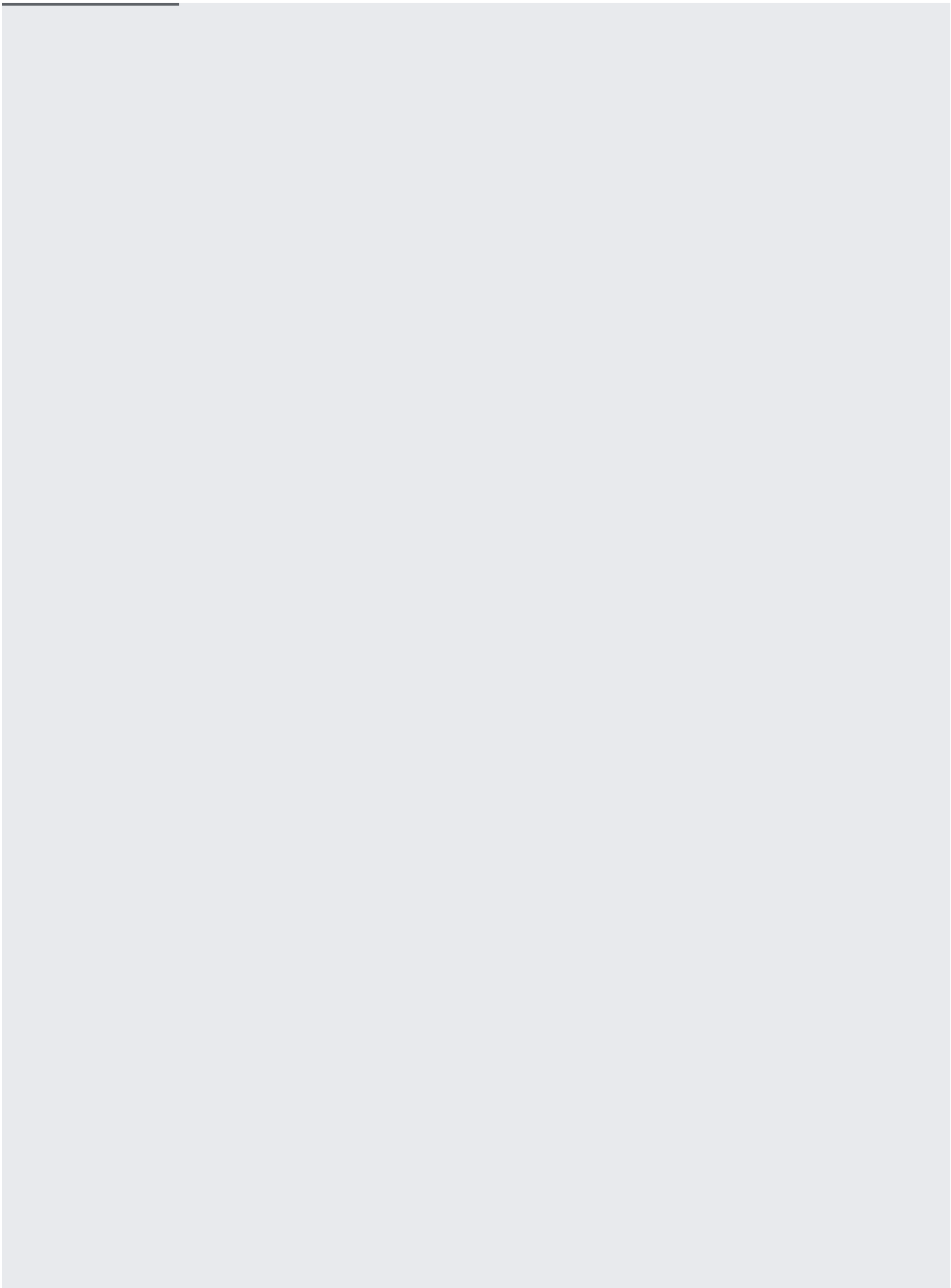


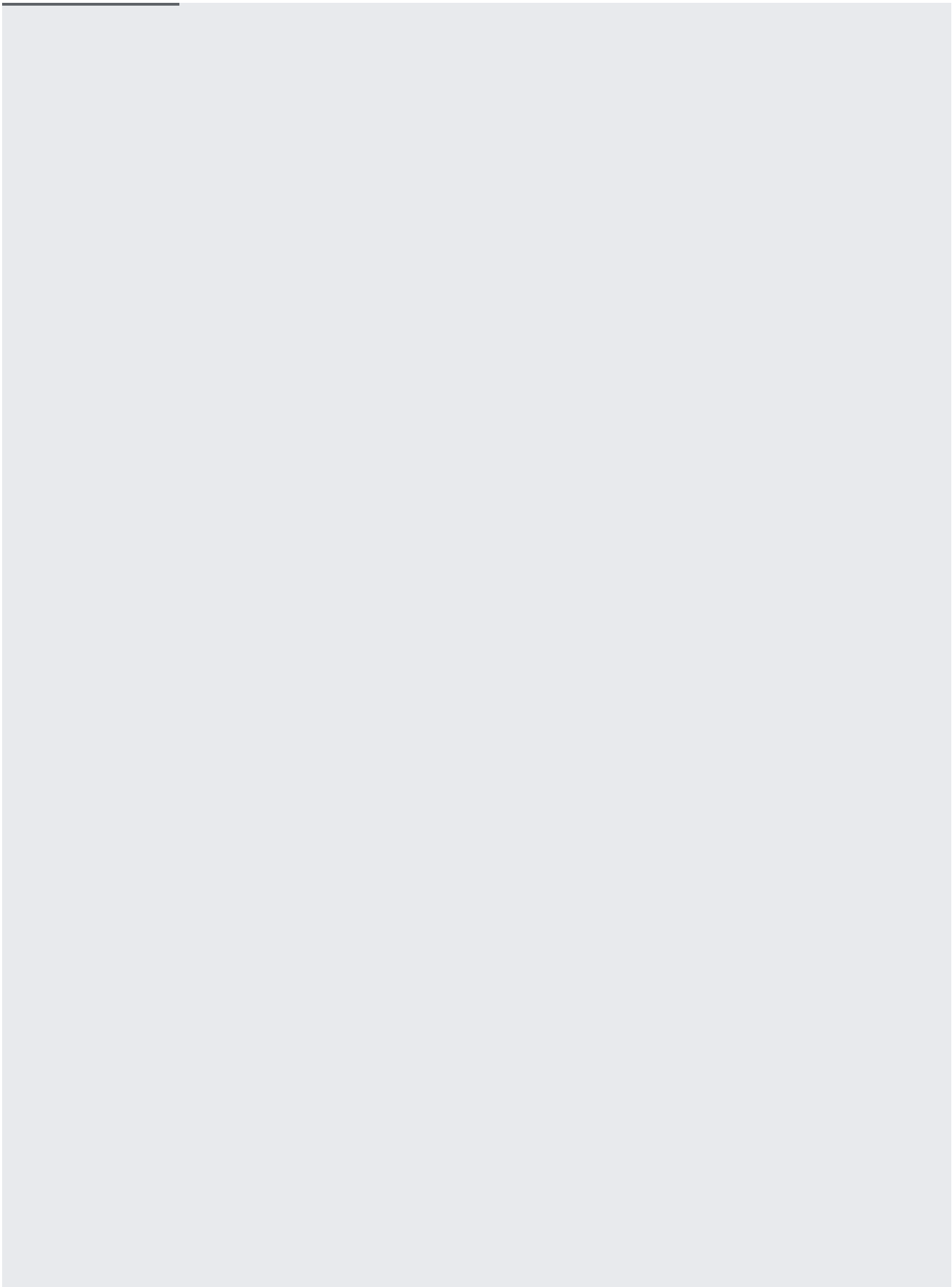




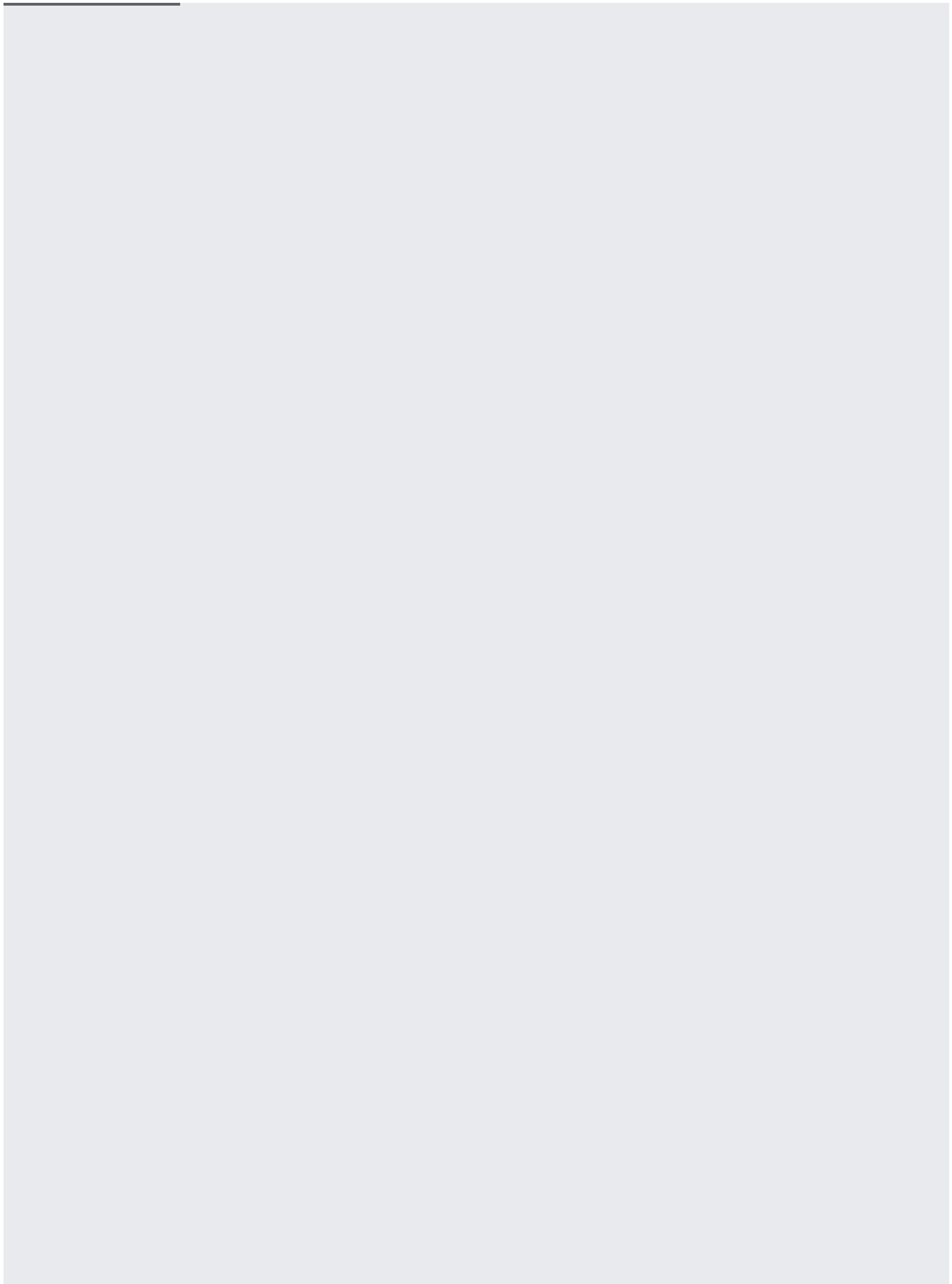
After you have created a dataset, you can import training sentence pairs into it. For details on preparing your training data, see [Preparing training data \(/translate/automl/docs/prepare\)](/translate/automl/docs/prepare).

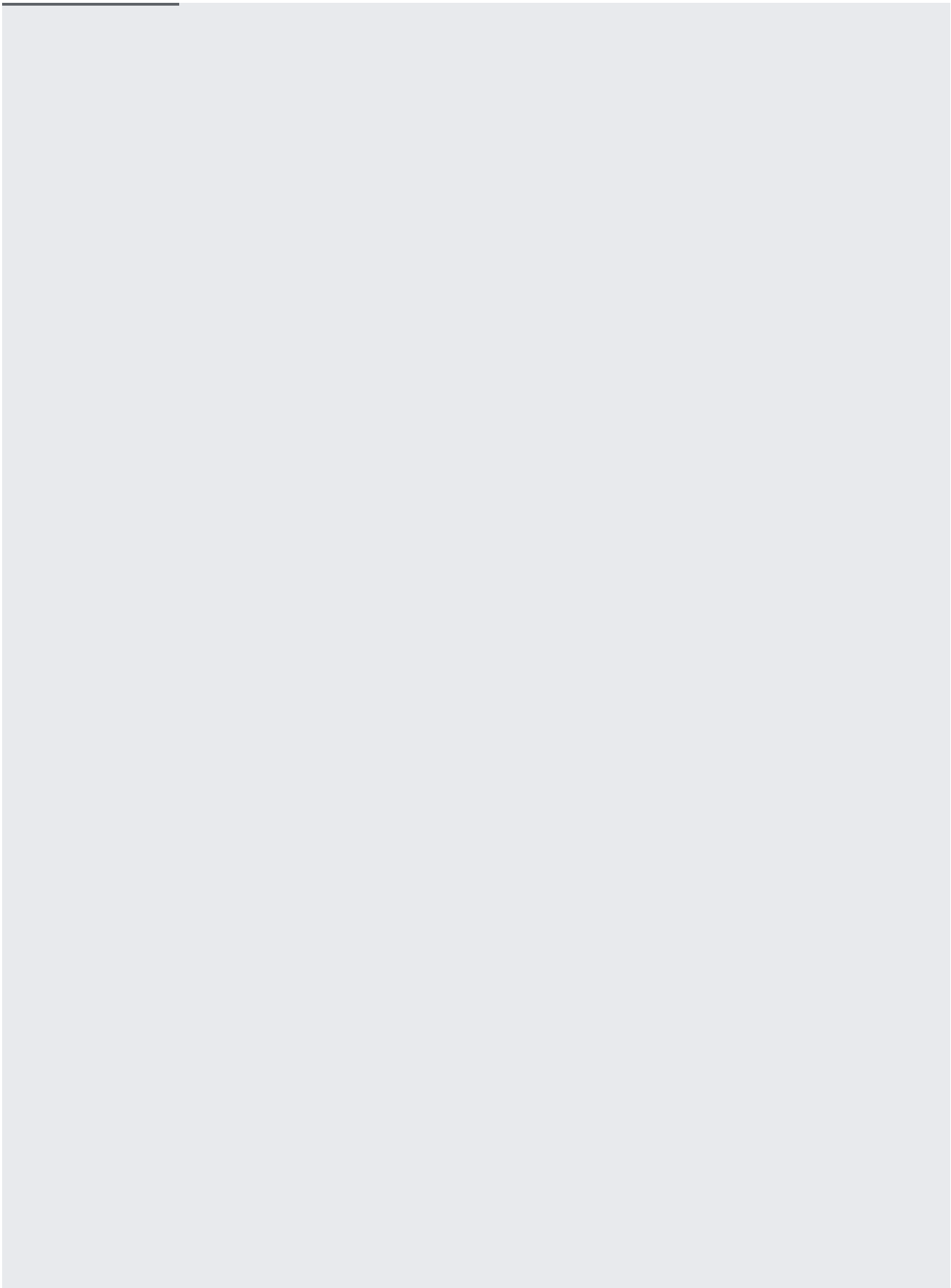


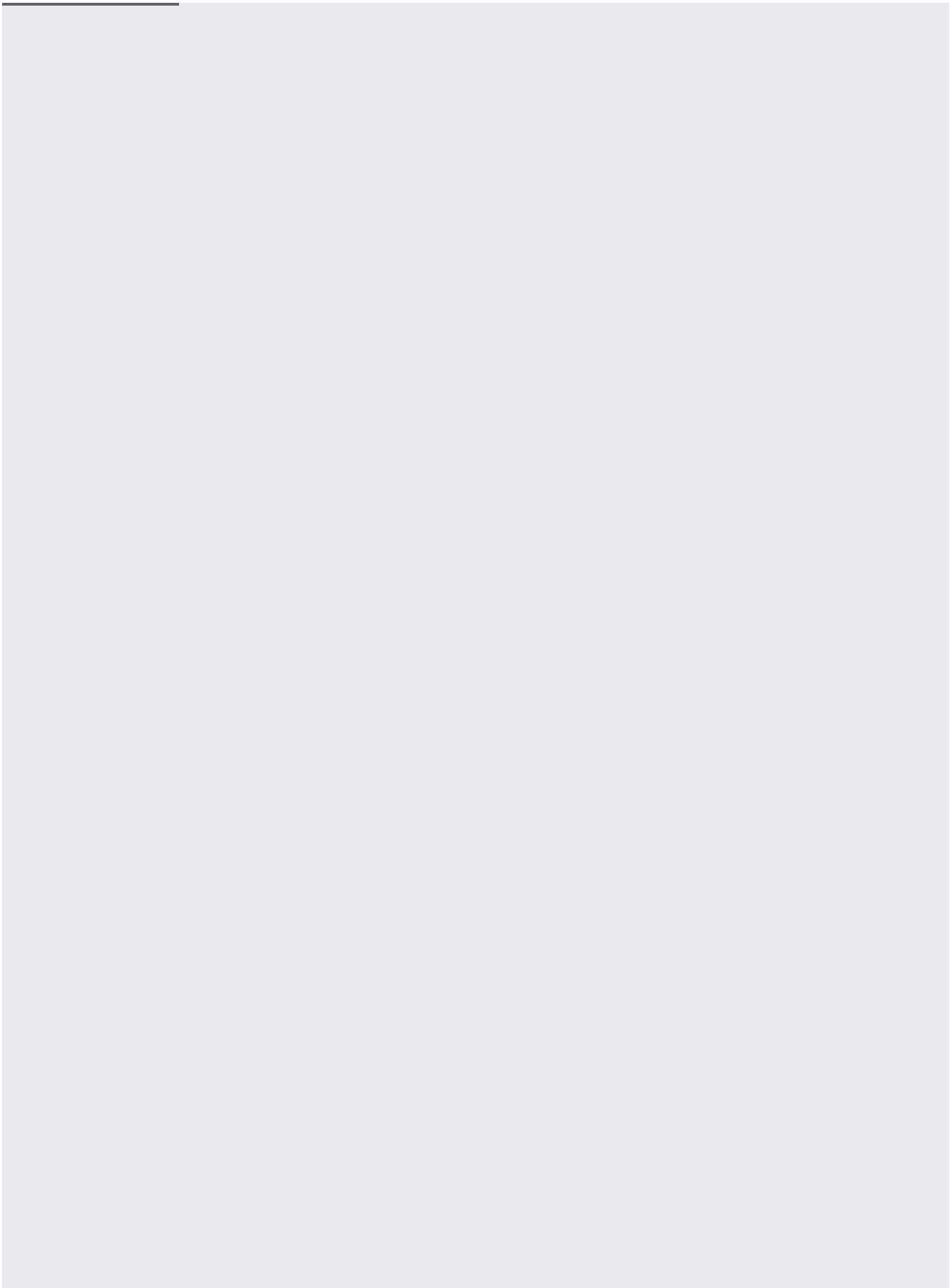


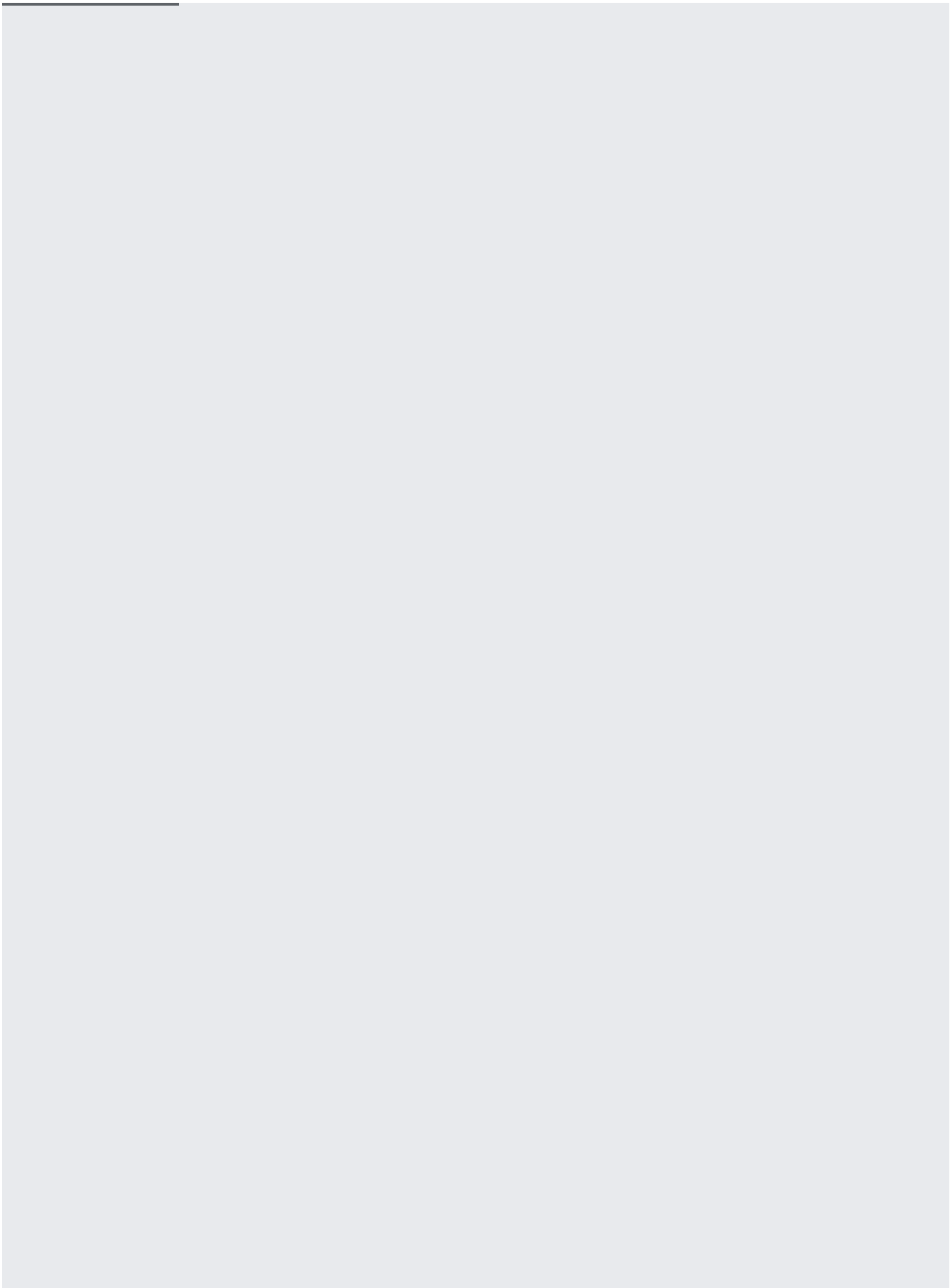


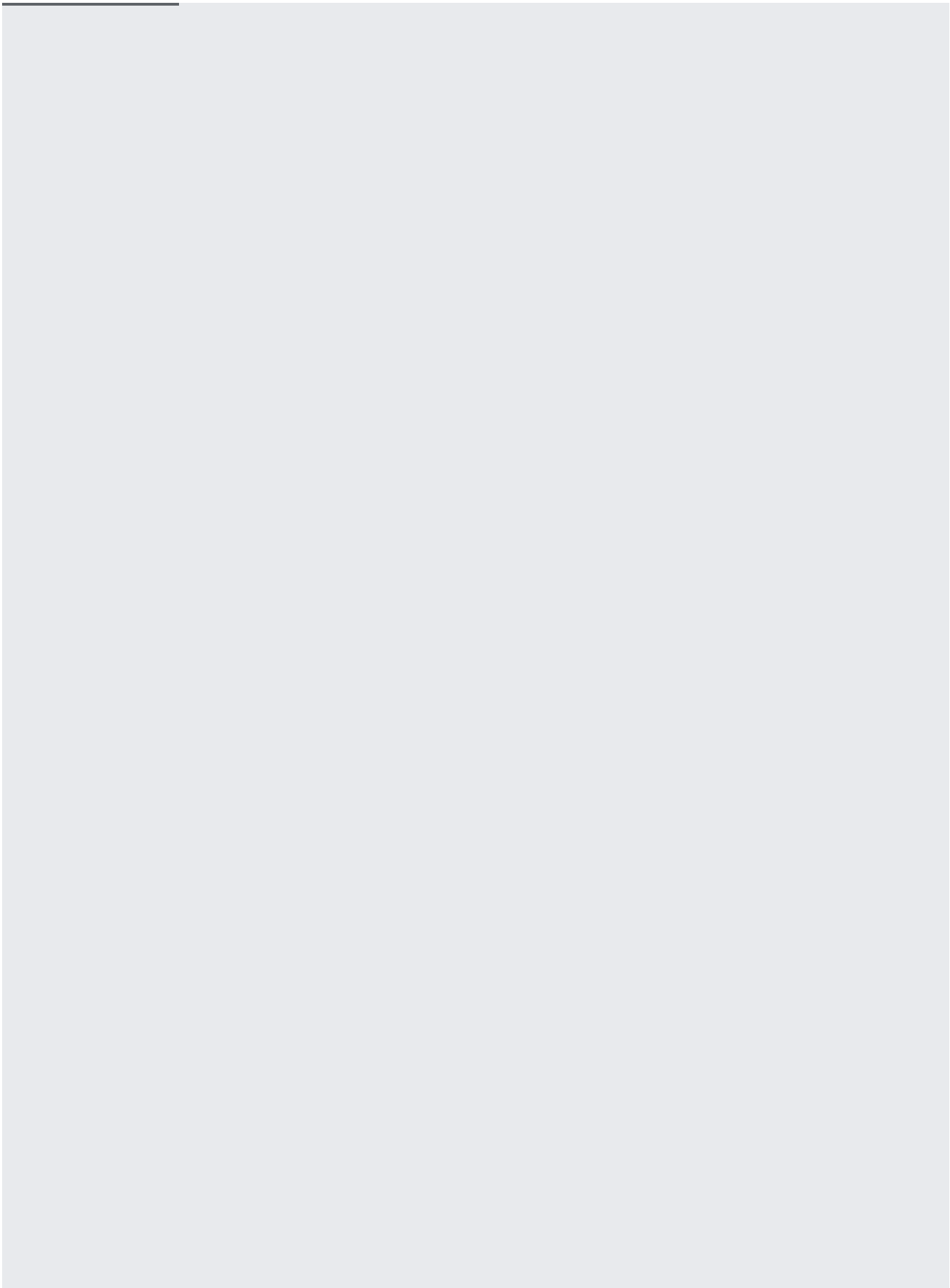


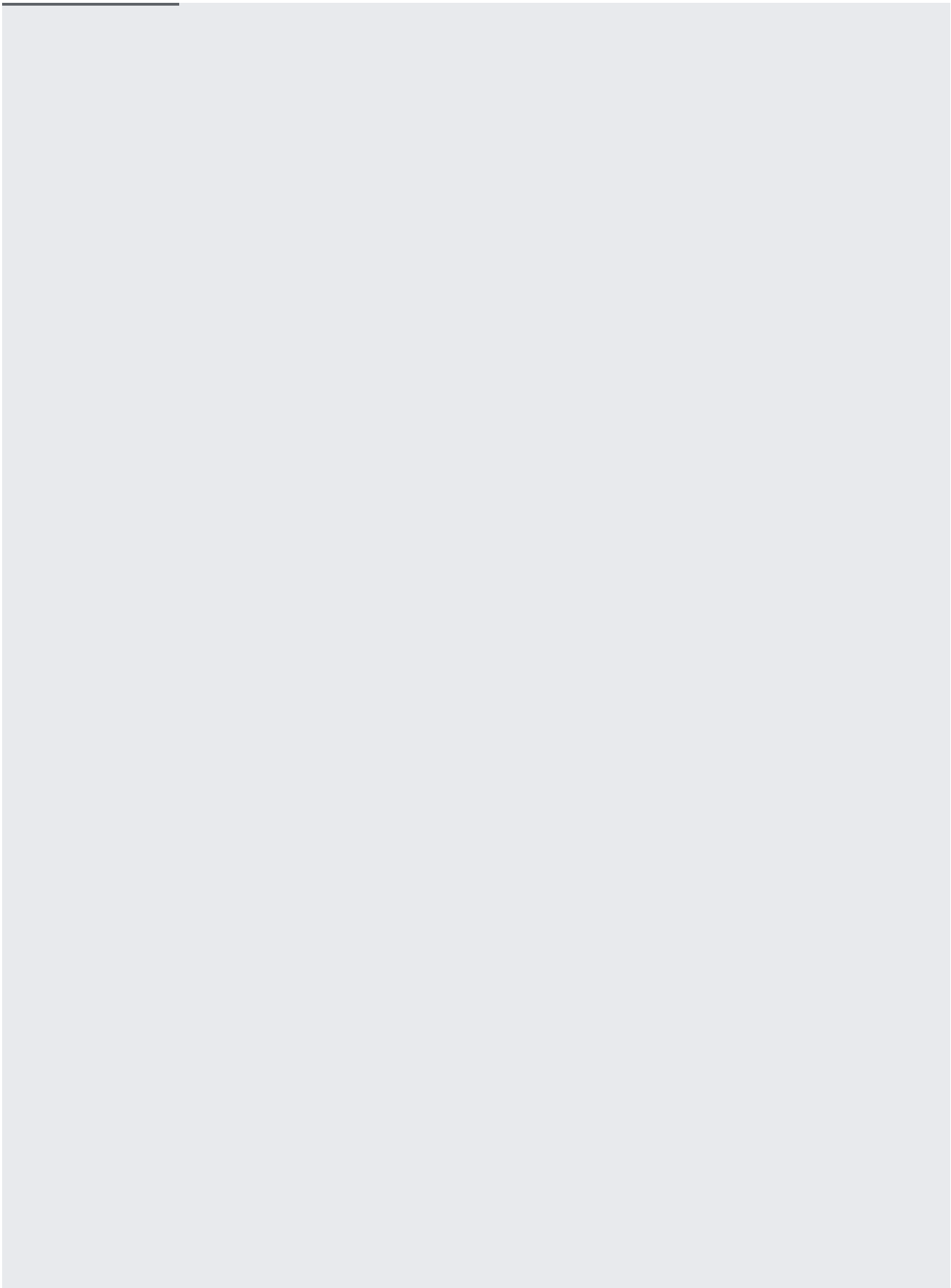












---

Once you have created and populated the dataset, you are ready to train the model (see: [Creating and managing models](/translate/automl/docs/models/) (/translate/automl/docs/models)).

A project can include numerous datasets. This section describes how to retrieve a list of the available datasets for a project.

