Name(s) _____ Period _____ Date _____

Activity Guide - Training Data

Activity: OpenAOverview

You are responsible for training a new large-language model to represent language. This requires large amounts of data that it can use to recognize important words and phrases. Your task is to curate this language data from your daily life so the language model can answer questions about the world around you.

☑ **Do This**: Complete the following tables with different pieces of text you would find in these locations. Include at most 5 words as examples of what a language model would learn from these sources.

For example: if your data source is "Street signs on the way to school", some words a model would learn could be "stop", "yield", "speed", "limit".

Curating Data

This Classroom	
Data Source (For example: books, posters, signs)	Words Learned from this Source (Write at most 5 words you might learn from a particular source)

Your Home	
Data Source (For example: magazines, instruction manuals, photos)	**Words Learned from this Source** (Write at most 5 words you might learn from a particular source)

On Your Way to School	
Data Source (For example: street signs, billboards, shops)	**Words Learned from this Source** (Write at most 5 words you might learn from a particular source)

Somewhere you go for fun (for example: the park, a mall, a sports arena, etc)	
Data Source	**Words Learned from this Source** (Write at most 5 words you might learn from a particular source)