

Learn more at [inspiritai.com](https://inspiritai.com) and email [emcmwong@stanford.edu](mailto:emcmwong@stanford.edu) with any questions  
Apply for our spring and summer programs at [inspiritai.com/apply](https://inspiritai.com/apply) by March 31

## Building AI Projects for High School Students

Daniela Ganelin, Inspirit AI Director of Curriculum

Many Inspirit AI students join the [AI Scholars Live Online](#) program because they want to apply AI to problems they care about. During the program, students work in small groups, guided by an expert [instructor](#), to complete an AI for Social Good [project](#). We also often hear from students who want the extra challenge of independently designing and completing their own projects! In this post, I'll walk through the process of creating your own machine learning project from scratch.

### Step 1: Practice machine learning basics.

Before you start your own project from scratch, it's a good idea to walk through some existing ones!

Find a tutorial using a resource like [Machine Learning Mastery](#), [Kaggle](#), or the [TensorFlow tutorials](#), and work through it line-by-line. [Google Colab](#) is a useful programming environment - it lets you jump right in to machine learning in Python with Google's hardware, without needing to install anything. Try to find a tutorial that matches your interests: simple numerical data is a great place to start, and later you could work on projects that explore natural language processing or computer vision.

Try to test yourself as you go! Explain your code to a friend; experiment with changing the code to see what happens; try to take the code to the next level by creating a new model or visualization. Get comfortable with looking up answers online - [Stack Overflow](#), [documentation](#), and your favorite search engine are your friends!

### Step 2: Choose a problem

In your (supervised) machine learning project, you'll train a model to make predictions from data. So what kind of problem do you want to solve? Maybe you're passionate about using AI to fight [climate change](#), so you want to build AI that can identify deforestation. Maybe you want to encourage kids to read, so you're building a [book recommendation system](#). Maybe you want to build a computer vision system to catch your dog sneaking into the treat bin!

Brainstorm some problems and [AI project ideas](#). If you're not sure, you can build off an existing tutorial or find a dataset that looks interesting and go from there!

### Step 3: Find your data

Next up, you'll need to find a dataset. In most domains, you can find lots of datasets online: [Kaggle](#), [data.gov](#), and the [UCI Machine Learning Repository](#) are a few great sources. You could also try emailing a researcher or blogger to see if they can share data.

If you're up for more of a challenge, you could collect your own data! This could involve using an [API](#) to pull data from a service like Twitter, [scraping](#) data from a website, or using data you've created yourself, like your personal photo library.

### Step 4: Prepare your data

Data pre-processing isn't glamorous, but it's important! Typically, you'll need to get your dataset uploaded to Colab, cleaned up, and represented in a nice tabular format (like a spreadsheet). [Pandas](#) is a helpful tool for handling and visualizing data. You might need to remove duplicate or missing data points, convert categorical data to numerical, and more. Spend some time exploring your data: can you make some [graphs](#) to explore general trends? Are there any outliers or oddities that jump out at you?

As you prepare for supervised machine learning, you'll need to [separate](#) your data into parts: you'll set aside your output (what you want to predict) from the input (the features that you'll base your predictions on). You'll usually [divide](#) your data into a *training* dataset that you learn from, a *validation* dataset that you'll use to evaluate your model as you improve it, and a *test* dataset that you'll use to measure the performance of your final model.

### Step 5: Build your model

Here's the part you've been waiting for: let's do some machine learning!

First, you'll need to decide what [sort of problem](#) you have: is this regression, where you're predicting a number, or classification, where you're predicting a category? Or maybe you're trying something more exotic, like [reinforcement learning](#) or [generative models](#)!

For now, let's focus on supervised learning. Using [Scikit-learn](#), you can train a model in just a couple of lines of code. You can start with a simple model like [linear regression](#) (for regression) or [logistic regression](#) (confusingly, for classification). You'll train your model on the training data and evaluate it on the validation data. If you're not happy with your initial results, you can experiment with trying other [models](#) or tweaking the hyperparameters (model options). Using [Keras](#), you can even put together a neural network in a few lines of code! (Beware: it's often tricky to find settings that work well.)

### Step 6: Evaluate your model

As you experiment, you'll need to evaluate your model on the validation set. Once you've settled on a model, you'll evaluate it on your test set to get a sense of its real-life importance.

You'll need to decide *how* you measure your model's performance! Sometimes this is as simple as measuring [accuracy](#). But often, measuring success and failure is more nuanced. Besides choosing the right [metric](#), you'll also need to think about some bigger-picture questions: is your model doing a good thing?

For example, would you prefer your medical AI to diagnose too aggressively (and tell healthy people they have a disease) or too conservatively (and miss people who are sick)? Can you check that your face recognition model [works well](#) for people of different genders and skin tones? Are you confident that your model will be used for good, or could it be used (intentionally or not) to [harm people](#)?

### Step 7: Deploy your model

Once you're feeling good about your project, you're ready to share and publish your work!

Colab notebooks are easy to share with friends and classmates. If you want to deploy your model more permanently, you could construct an [app](#) or website for it: we're fans of a quick new web app creation tool called [Streamlit](#). Be sure to include information about the process you used to create your model and its limitations, so that users can decide how they want to use it!

To disseminate your work more broadly, you could try sharing it in online forums or submitting to a [competition](#) or local science fair. You could even try reaching out to professors or researchers to ask if they could serve as project mentors: maybe your work could become a publication one day!

For inspiration, check out some incredible machine learning and data science projects that teenagers have completed with the aid of mentors: [Kavya Kopparapu's](#) app to diagnose eye disease, [Lillian Petersen's](#) tools to predict crop harvests to fight hunger, and [Raghav Ganesh's](#) device to help visually impaired people spot obstacles. Today, Raghav is an instructor at Inspirit AI!