Principles in Action Playbook: Development

How do you build something with AI?

- □ Do you need to build your own AI?
- □ Where should you get your data from and how do you evaluate it?
- □ How do you build a responsible model?

Do you need to build your own AI?

Building AI solutions differs from traditional software development where there are often defined product milestones, requirements, and estimates.

Whether or not you decide to build your own AI depends on various factors such as your specific **goals, resources, expertise**, and the availability of **suitable AI solutions** on the market.

Here are some reasons to build your own model:	**Here are some reasons not to build your own model:**
□ You and your customers need to understand exactly why something happened □ You have access to high quality or custom data for training □ Full data transparency is needed □ You have the capacity, desire, and willingness to adhere to responsible AI principles and regulatory requirements	□ Investing in intensive research and exploration is not a priority □ Sourcing sufficient high quality data for model training and testing will be challenging □ Your team does not have the expertise and resources in machine learning, data science, and software engineering □ You do not have the budget to support infrastructure costs and ongoing maintenance

Where should you get your data from and how do you evaluate it?

If your team has decided to build its own AI model or fine-tune an off-the-shelf solution you will need data for training and testing. **Your AI model**, and thus your product, **will only be as good as the data and labels that feed it**, so think through your data needs carefully.

Consulting subject matter experts will greatly help you in this process. Domain experts don't need to be data experts, they just need to be willing to share insights and highlight implications about your data's subject matter.

Let's consider how we source our dataset:

Do you have a data card? What information does it document about your dataset?

Do you require a licence to use the dataset or do you need to cite the publisher of the dataset? Are you legally allowed to use and store the data in the geographic regions you plan to release your product?

What preprocessing has the data gone through?

How closely is the dataset representative of your users and your use case?

- □ Does the data reflect the real world? Consider user demographics, recency, time of year, trends, global events, image quality, and mistakes in text.
- □ Is the data noisy? That is not necessarily a negative; allowing for imperfect data can more closely match the data you will get from your user base.
- □ Do you need to modify the dataset with additional data or augmentation techniques, or combine multiple datasets?
- □ If you can't find representative data, are you comfortable limiting your product release to only demographics of users reflected in your dataset?

Does the data source have known biases? What are they?

What is the quality of your data labelling?

- □ Can you trust your labellers and labelling tools to sufficiently and accurately label data with minimal bias?
- □ Have you given the labellers sufficient guidelines to label and agree on labels?
- □ Have you set up a labelling process that compensates labellers fairly, ensures safe working conditions, and respects workers' ethical boundaries on sensitive data?
- □ Should you train domain experts to create gold standard labels?
- □ Are there mistakes? Scrub the data for missing values, duplicates, inconsistent formatting, or incorrect labels.

Is the dataset adhering to privacy and security standards?

- □ Have you redacted all personal identifiable information?
- □ Have you aggregated data to maintain anonymity?
- □ Do the right people have permission to access the data securely?
- □ Is data encrypted and stored securely?
- □ Have you considered privacy methodologies - including differential privacy, federated learning, homomorphic encryption, or synthetic data generation – to address privacy risks and mitigate them?

How will you maintain your dataset going forward?

- □ Do you have manual or automated data inspection and quality assessment mechanisms to ensure the quality of data?

  How will you know when data is outdated?

How do you build a responsible model?

Now that you have good quality data that reflects your users and use case, you can start thinking about how to develop the model that will output predictions or content that will help address your **users' needs**.

Are you using an off-the-shelf model? How stable is it? Most users aren't concerned about which state of the art model you're using, only that they are getting the information they need to get their task done.

What societal biases, both explicit and implicit, might influence your team's decisions during model development? Let's acknowledge them.

Does your model's output impact human well-being, such as healthcare, employment, justice, or finance? How will you prioritise explainability and interpretability of your model?

Are you training the model to be robust against adversarial attacks? How?

□ Is your model trained on a secure network to protect data and access?

□ Are you aware of your carbon footprint as a result of training, maintaining, and running inference on your machine learning models?

The idea of achieving an optimally responsible model is a fallacy; developing models is a **constant balance of making trade-offs for your unique use case**.