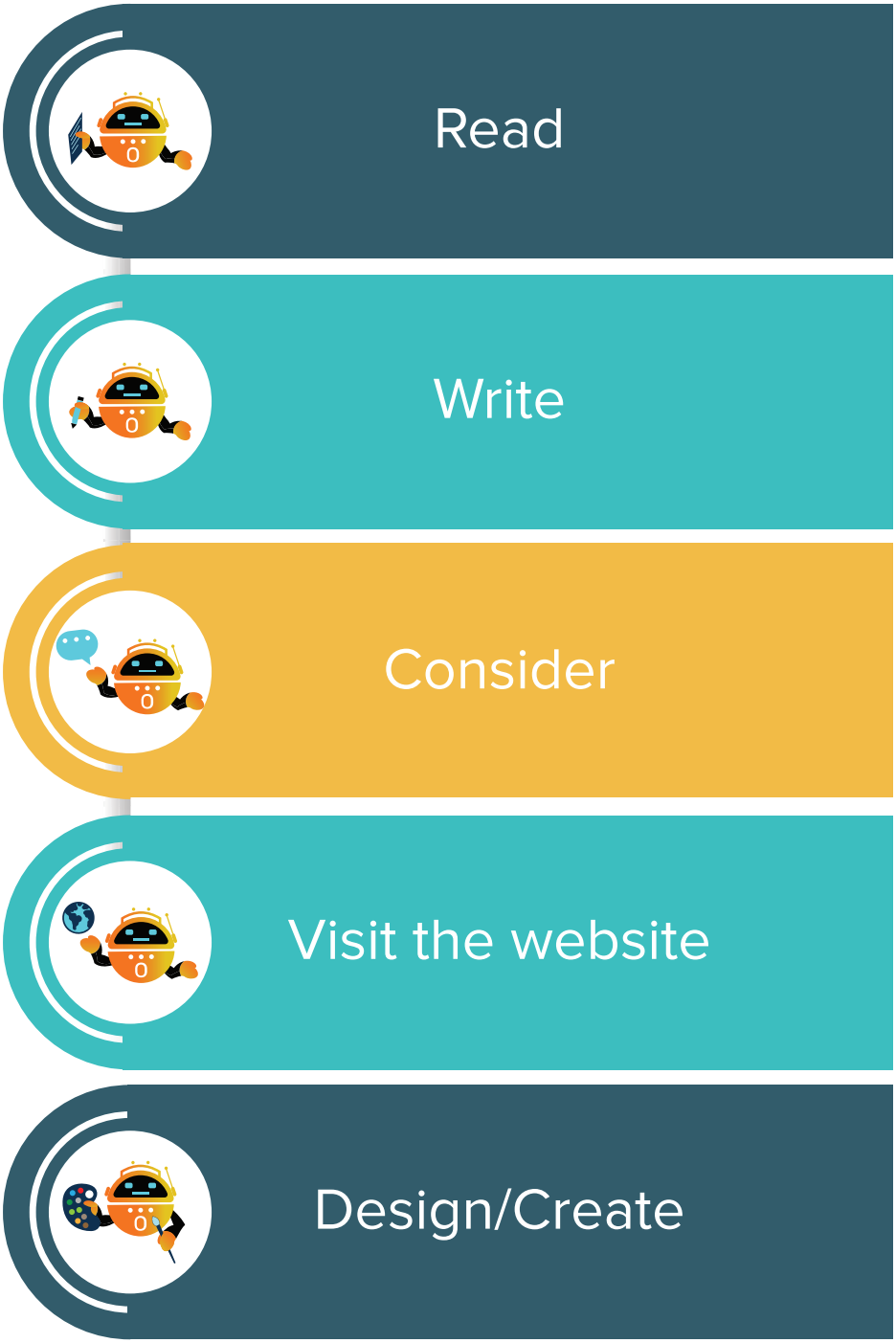
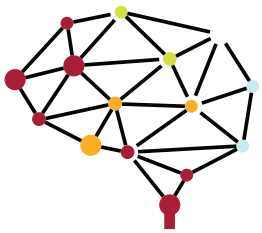
xPRO

Designing and Building AI Products and Services Workbook

Your AI Mascot will guide you through this week, so watch out for these icons:



Workbook 8: Final Capstone Project



Week 8

Design and develop a product using AI with the aid of the resources you have learned about throughout the course. The application you decide to develop in your project is completely arbitrary; it can be something that has to do with your interest or hobbies or something that is closely connected with your educational and/or working background. To help you, we have listed the main points below or parts that will help you throughout each phase of your design project.

1 Step One: AI Design Practices

As you have learned in Module 8, AI design practices involve four stages, namely boundary setting, value formation resulting from explanation, and bias remediation. Briefly explain these four steps.

AI systems have come under scrutiny lately because of potential bias issues. Answer the following scenarios in terms of model bias.

Scenario One:

Assume that you implement a resume filtering tool using machine learning models. However, your model starts rejecting applicants based on their gender. What is the potential reason(s) why your model biases against a particular gender, and how can you solve it?

Type your response here

Issue :

The main reason for model not recognising resumes under all Gender is due to “ Class Imbalance Issue ”. The model has been trained on set of samples with minority population of one gender.

This could be due to various reasons:

1) Lack of equal proportion of gender records due to dataset being Small.
2) Similar to example of Anamoly identification, it could have been the case where the resume with minority gender would be less available or missing completely.
3) Scope of the model would have been defined to restrict under One specific gender.

EDA:

To understand the problem more clearly, its necessary to perform exploration of data by few simple 2 dimensional plots eg: #volume of records over gender or with regard to other demographics.

Solution:

Augment Dataset using : “ Under Sampling ” , “ Boosting ” or Reviewing Scope.

1 Step One: AI Design Practices(Cont.)

Scenario Two:

Assume that you want to detect attack samples over the internet by implementing an AI system. However, your training dataset is imbalanced where it consists of a small number of attack instances. In such circumstances, it is challenging to classify samples as an attack or non-attack using traditional machine learning algorithms since state-of-the-art machine learning models cannot learn the characteristic behavior of the minority attack class. As a result, models are easily biased to the majority class. One naïve solution is to remove the necessary amount of non-attack samples and make the dataset balanced with a small size. What is the limitation of this solution? How can the GAN idea be integrated into this imbalanced training dataset issue to produce a more efficient solution?

Type your response here

Issues by applying “ UnderSampling ” could be , the model would not have understood and trained to identify malignant defects. It needs necessary features to classify between defect over good samples. So datasets must be decent propotional to understand between good and bad.

This can be mitigated by leveraging GAN with appropriate augmentation of dataset. Using generator of GAN, we need to create various cateogories of defect samples. So we need to take the help of SME ’ s or System experts to understand what are the various forms of Bad samples and ensure GAN can generate those defects datasets.

So integration of GAN to augment the dataset will helps to create large or required volume of datasets so to handle Imbalanced training dataset issue.

2 Step Two: Transfer Learning and Online Learning

Assumption One:

Assume that you want to create an AI system to improve production of your company. However, you don't have enough training samples to implement an AI system. You propose using a pre-trained model to reduce the computational demands of learning. Such an approach is often associated with the idea of transfer learning. Provide a thorough description of transfer learning and how and why it works including a concrete example?

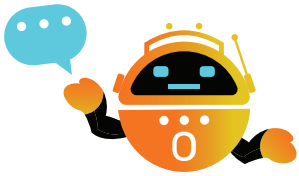
Type your response here

Definition:
Leveraging general purpose or domain specific ML model that ' s pretrained with huge dataset and augment with required specific set of model or networks to restrict for a classification is called " Transfer Learning " .

Process:
As the early layers of DNN involve with Feature extraction which is going to be common for images, the transfer learning helps us to utilise learnings gained from these layers. For instance, the weights gained from pre-trained model can be used rather than random weights. The pretrained base model can be instantiated based on architectures such as ResNet or Xception. However, one or more final layers needs to be replaced with problem specific fully connected DNN network with required hyper parameter values.

When we need this:
This is especially needed while dealing with small dataset or under resource constraints. The best use cases are for NLP and Computer Vision.

Specific Example:
To recognise



2 Step Two: Transfer Learning and Online Learning(Cont.)

Assumption Two:

Assume that you generate an AI system to create a fully automated model for your company's needs. Your model initially works fine as it is. However, your company produces new training data daily that is rapidly or slightly different from your initial training samples. Therefore, your AI model is not suitable to solve the problem for new samples in a dynamic setting. Provide a thorough description of online learning and how and why it solves the problem under the a forementioned circumstances.

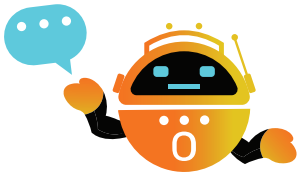
Type your response here

Purpose:
Online Learning is a technique that requires for an ML model to continously learn from its data and adapt parameters iteratively based on the data. This can be acheived by operationalising ML model via Automated pipelines.

Though a model has been trained based on set of parameters, the variations of data over time is very important for providing up-to-date predictions. For example: in News Industry, prediction even if based on yesterday ' s data provide stale results. Thus online learning is quite important for these models to get trained on new data as and when it arrives.

As it learns from its new data without retraining the entire dataset, this incremental learning process could be fast allowing it to provide real-time predictions. This is specifically needed for Autonomous Vechicle where detecting the obstacles in near-real time and takes decisions accordingly.

It ' s a more crucial concept especially while productionazing a ML model. Though this can be closely related to RealTime Learning, the main idea is about Iteratively update the model with latest data and avoid providing stale results to end users.



3 Step Three: Develop A Plan

This question consists of various sub-steps.

Part One:

Choose a problem that you want to solve with the help of AI. As an AI-based model solution to your problem, you need to think about the data-gathering efforts, infrastructure, storage, ETL, and cleaning you may want to conduct.

Think about what your project may require to complete this step. You can also search various resources like research papers and other publications that might help you in this process.

Type your response here

Please find attached

3 Step Three: Develop A Plan (Cont.)

Part Two:

This step involves incorporating the higher steps in the pyramid that we have defined as the AI hierarchy of needs. This step involves, among other actions, employing data analytics, aggregating the data, and training the data. Other parts of this step involve running A/B testing and simple ML algorithms to understand what to expect from your data. Finally, you will need to think about which AI deep learning method to apply to achieve your goals. Depending on your project, this step may also include designing an HCI.

Now, write a short paragraph about the second part.

Type your response here

3 Step Three: Develop A Plan (Cont.)

Part Three:

Another fundamental step when designing an AI product involves thinking about the humans in the loop, not in terms of a gathering of data, as you have described in the previous step—but more as a way of thinking about how humans may be involved in decision making (think about the example presented in Week 7 about autonomous vehicles).

Describe the number of ways you think that humans and your idea will be interacting and document them in a point-wise manner.

Type your response here

4 Step Four: Facing Possible Challenges

The last step of designing an AI product may involve thinking about facing the possible challenges, solving the problems, and extending the ongoing research to improve your product.

Describe the points mentioned above based on the AI product you want to design.

Type your response here

Congratulations on designing your first AI product!

MIT xPRO

Designing Artificial Intelligence Products Workbook