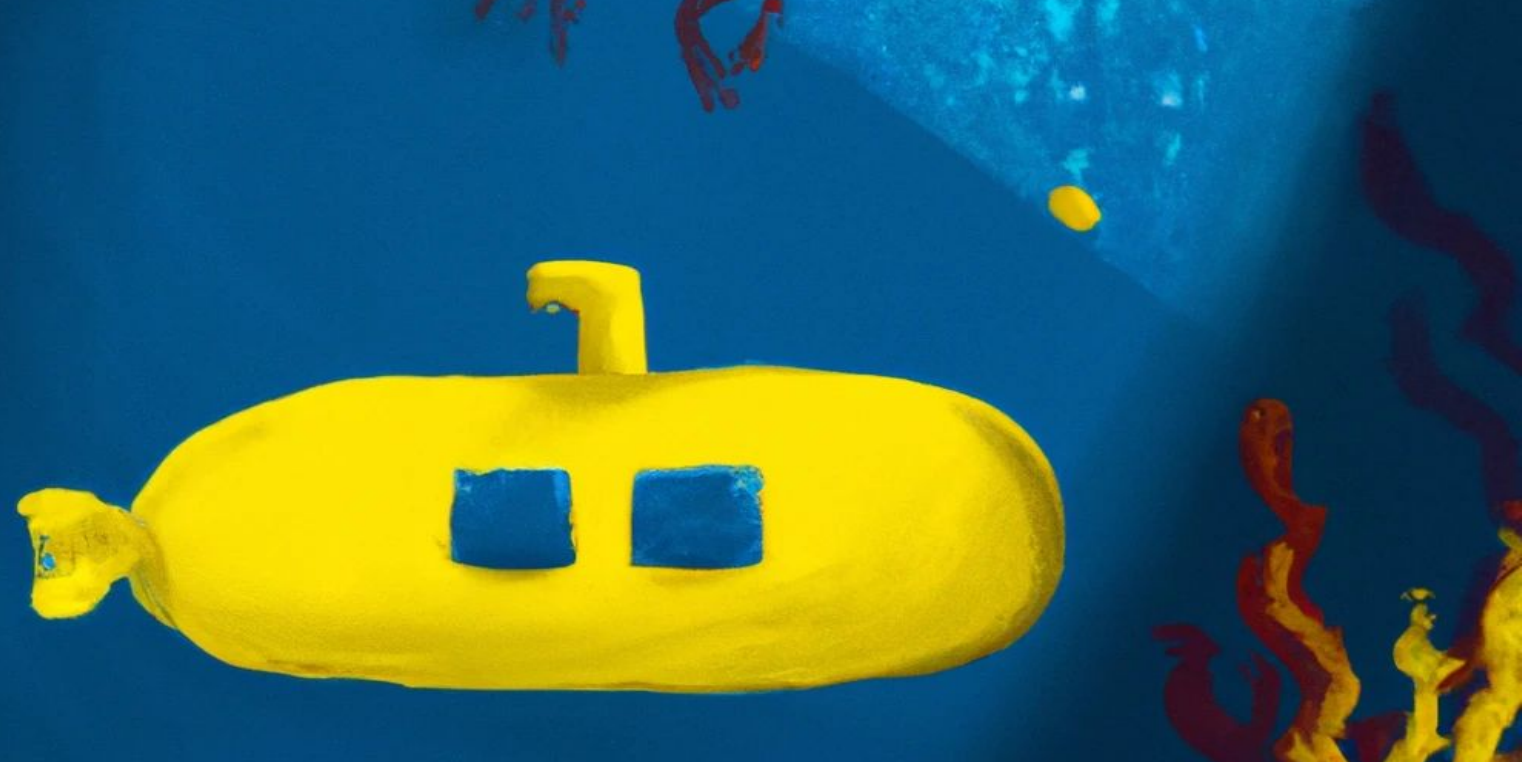
No class, watch recording in Media Folder (Canvas)

CSCI 1470/2470
Spring 2023

Ritambhara Singh

February 15, 2023
Wednesday

Deep Learning



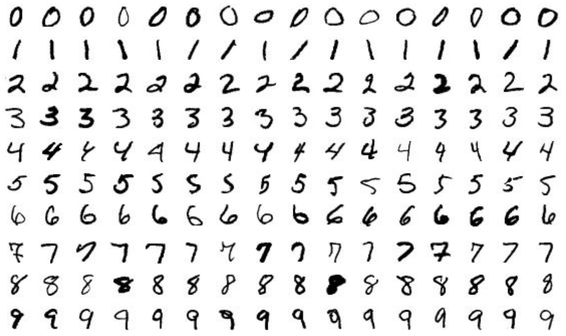
DALL-E 2 prompt “a painting of deep underwater with a yellow submarine in the bottom right corner”

Recap

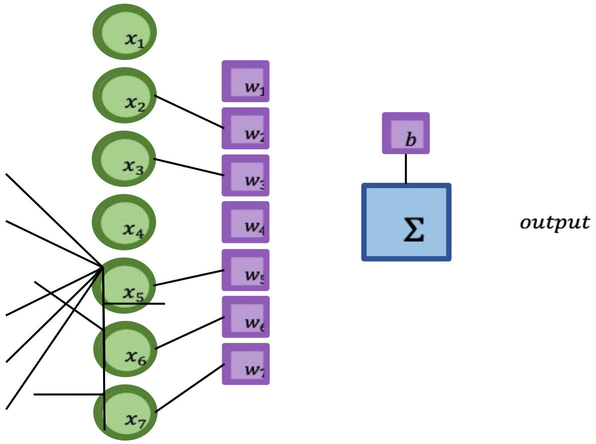
Intro to machine learning

Supervised Learning

Handwritten digit recognition task



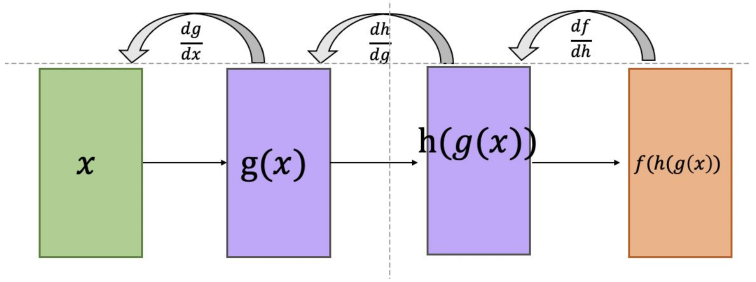
MNIST dataset



Perceptron and it’s learning algorithm

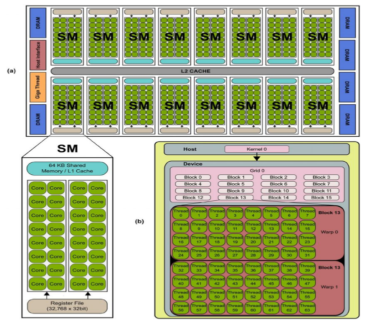
Loss functions

Building simple neural networks



Gradient descent and backpropagation

Matrix formulations and GPUs



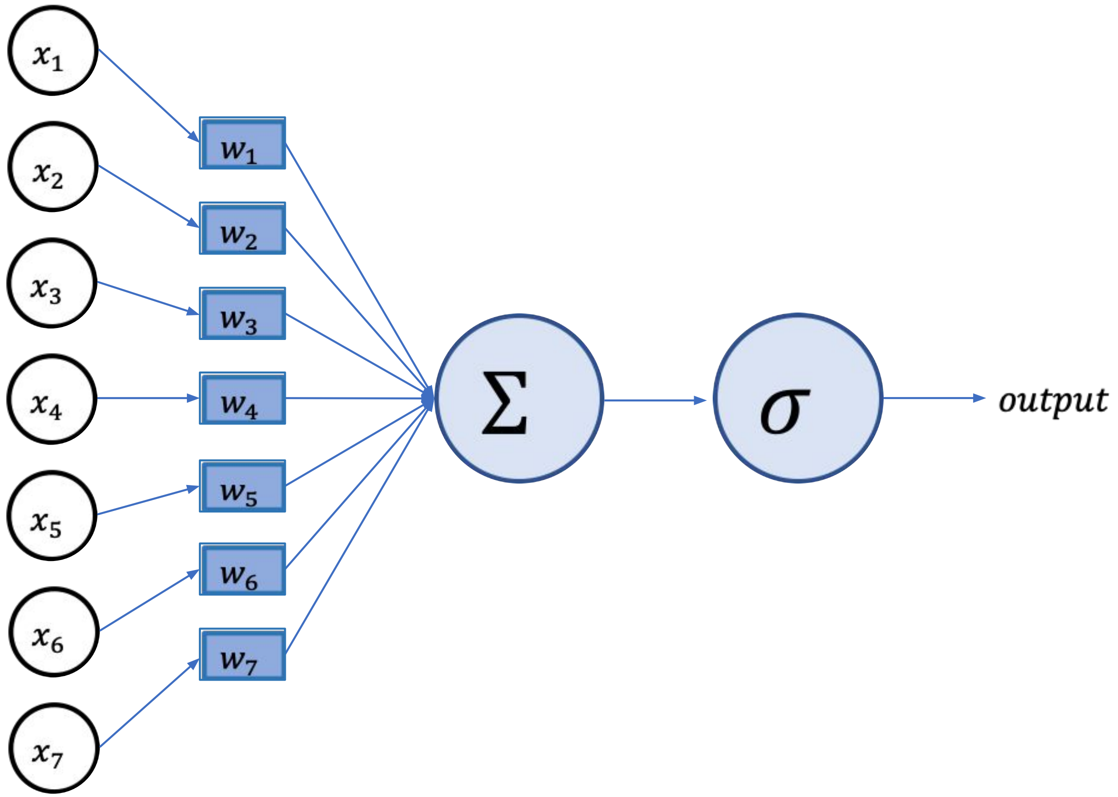
Today’s goal – lean about the lifecyle of machine learning systems

(1) What are the different stages?

(2) What are the different considerations?

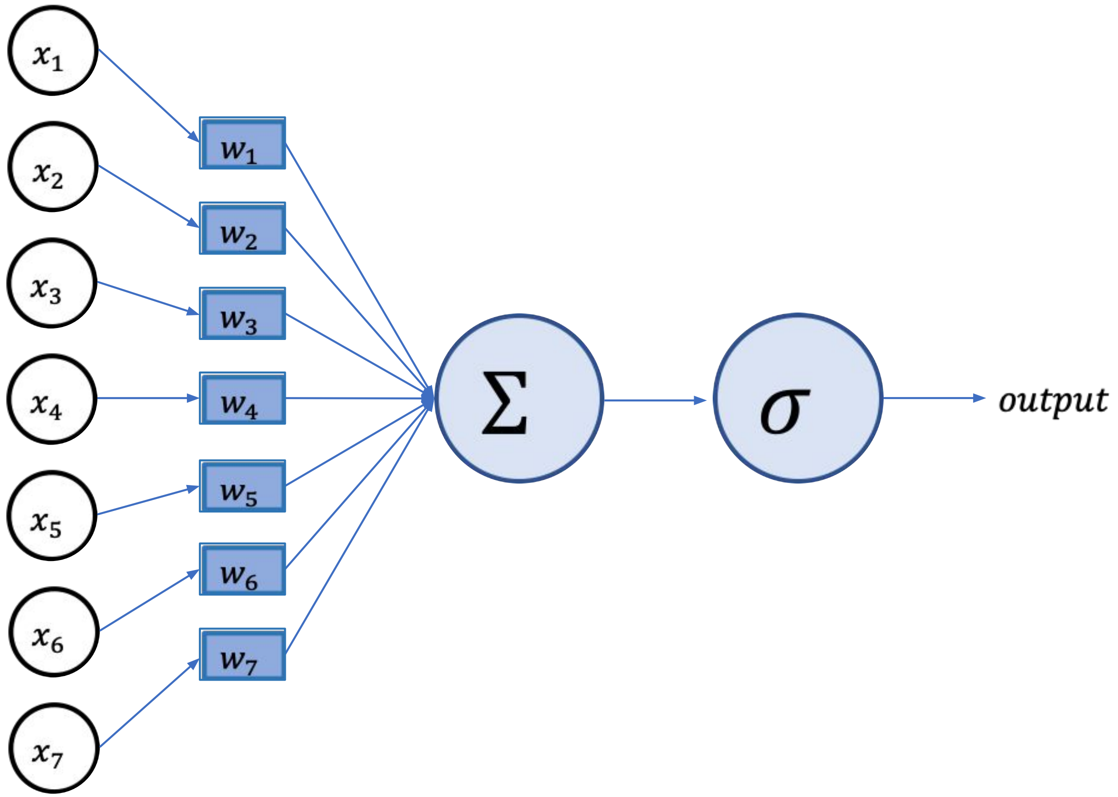
(3) Designing your Deep Learning system

The Primary Focus of This Course



The math behind deep learning

The Primary Focus of This Course



The math behind deep learning

```
def main():
        train_input, train_labels = get_training_data()
        test_input, test_labels = get_testing_data()
        model = Model()
        optimizer = tf.keras.optimizers.SGD(learning_rate=1)
        for i in range(num_epochs):
                train(model, optimizer, train_input, train_labels)
                print("Epoch: ", i)
                sum_acc = test(model, test_input, test_labels)
                print("Test Accuracy: %r" % (sum_acc/100))
        return


if __name__ == '__main__':
        main()
```

How to implement specific models

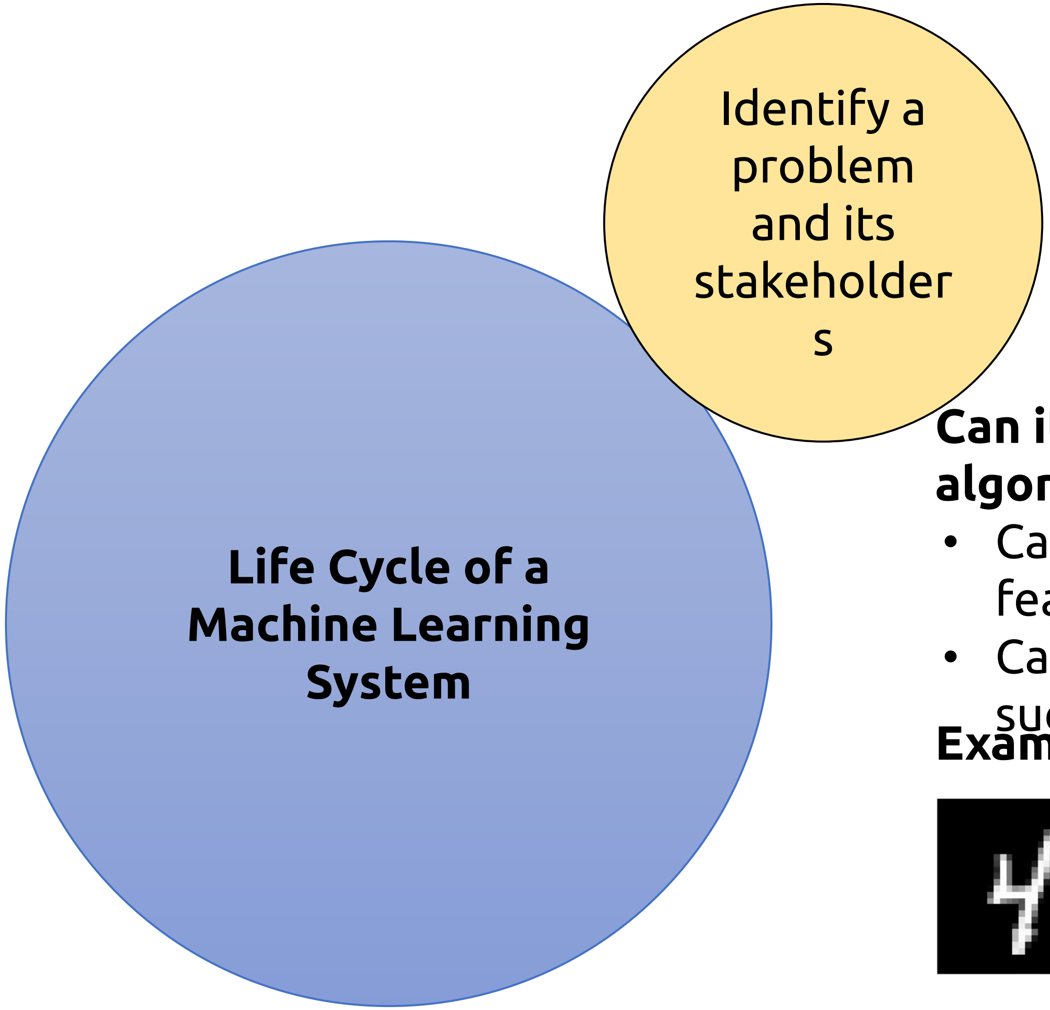
But how do our models *interact with the world?*

- DL/ML/AI systems are never “just math”
  - Developing a system is a lot messier than that, and there are not always clear “right answers”

The Life Cycle of Machine Learning Systems

Or, a framework for thinking critically about the impacts of the ML models that you develop

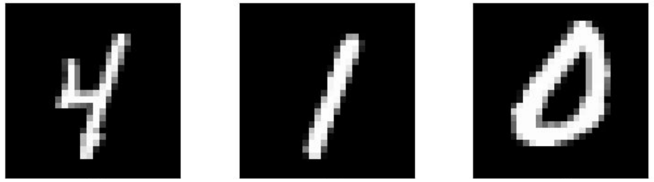
Life Cycle of a Machine Learning System



Can it be solved by an algorithm?

- Can we encode all relevant features?
- Can we define a metric for success?

Example: digit recognition



- *Features*: image pixels
- *Success*: test-set accuracy
- Seems like a clear **yes**

Identify a problem and its stakeholders

Life Cycle of a Machine Learning System

Can it be solved by an algorithm?

- Can we encode all relevant features?
- Can we define a metric for success?

Example: predicting online virality

- *Features*: ...the state of the world?
- Might be a **no**

Identify a problem and its stakeholders

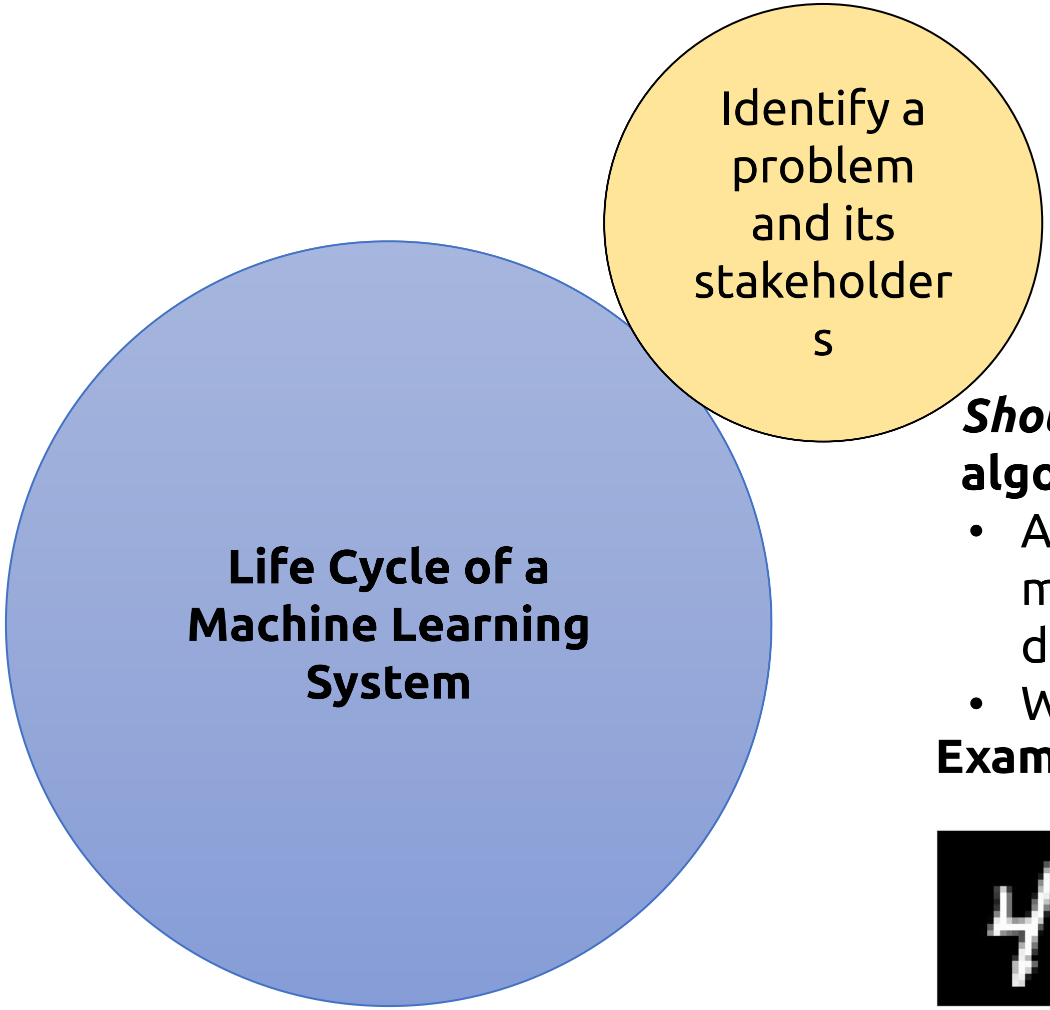
Life Cycle of a Machine Learning System

Can it be solved by an algorithm?

- Can we encode all relevant features?
- Can we define a metric for success?

Example: predicting effectiveness of classroom interventions

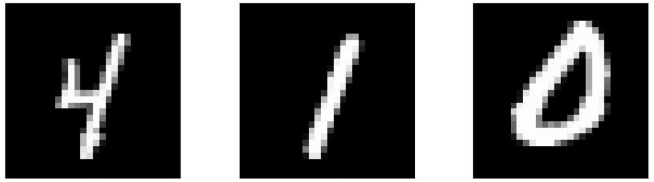
- *Success*: different stakeholders (teachers, parents, administrators, govt. officials) may not agree…
- Might be a **no**



Should **it be solved by an algorithm?**

- Are we comfortable with a machine making this decision for us?
- Who is “we” in the above?

Example: digit recognition



- Leads to faster mail sorting, which is probably a good thing
- Seems like a clear **yes**

Identify a problem and its stakeholders

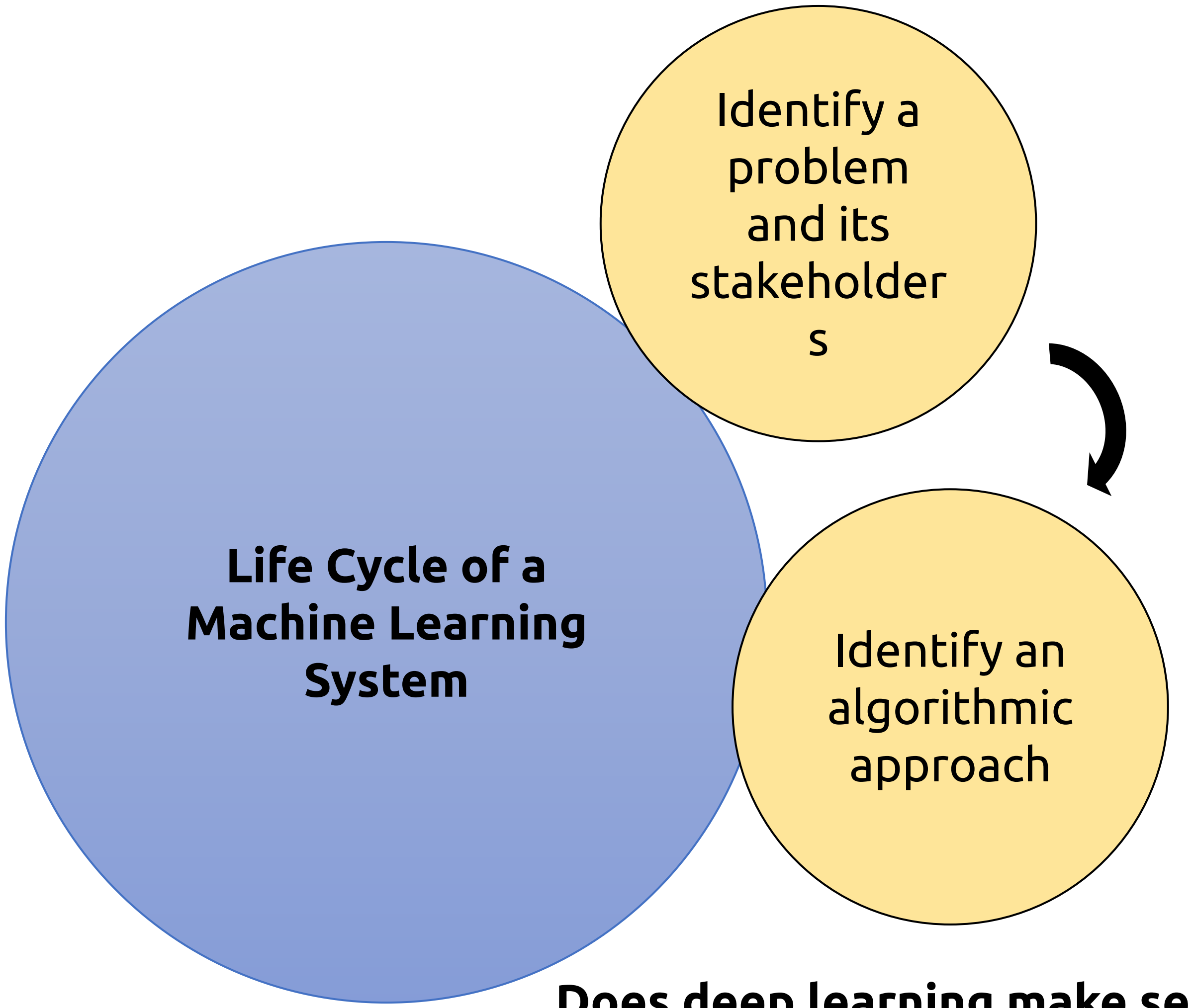
Life Cycle of a Machine Learning System

***Should* it be solved by an algorithm?**

- Are we comfortable with a machine making this decision for us?
- Who is "we" in the above?

Example: autonomous w

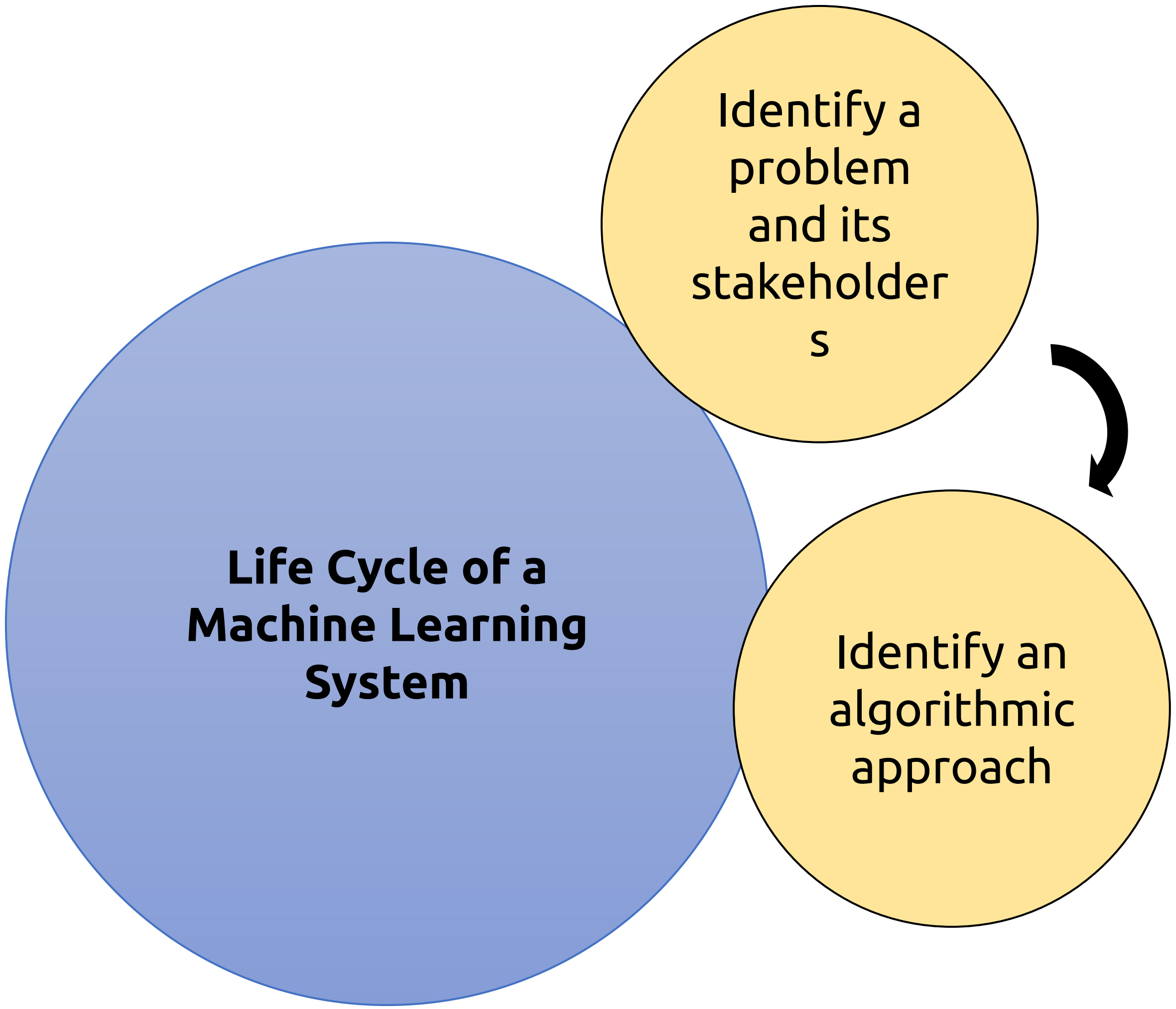
- Drones and robots getting involved in combat?
- Thousands of AI researchers



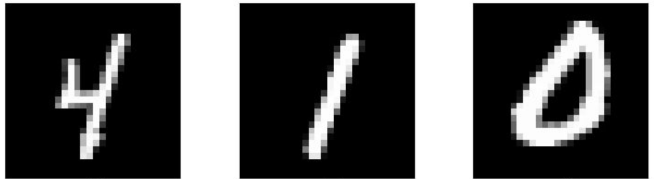
Does deep learning make sense?

- Can we acquire enough training data?
- Are we ok with not understanding why the model makes the decisions it does?

Model interpretability



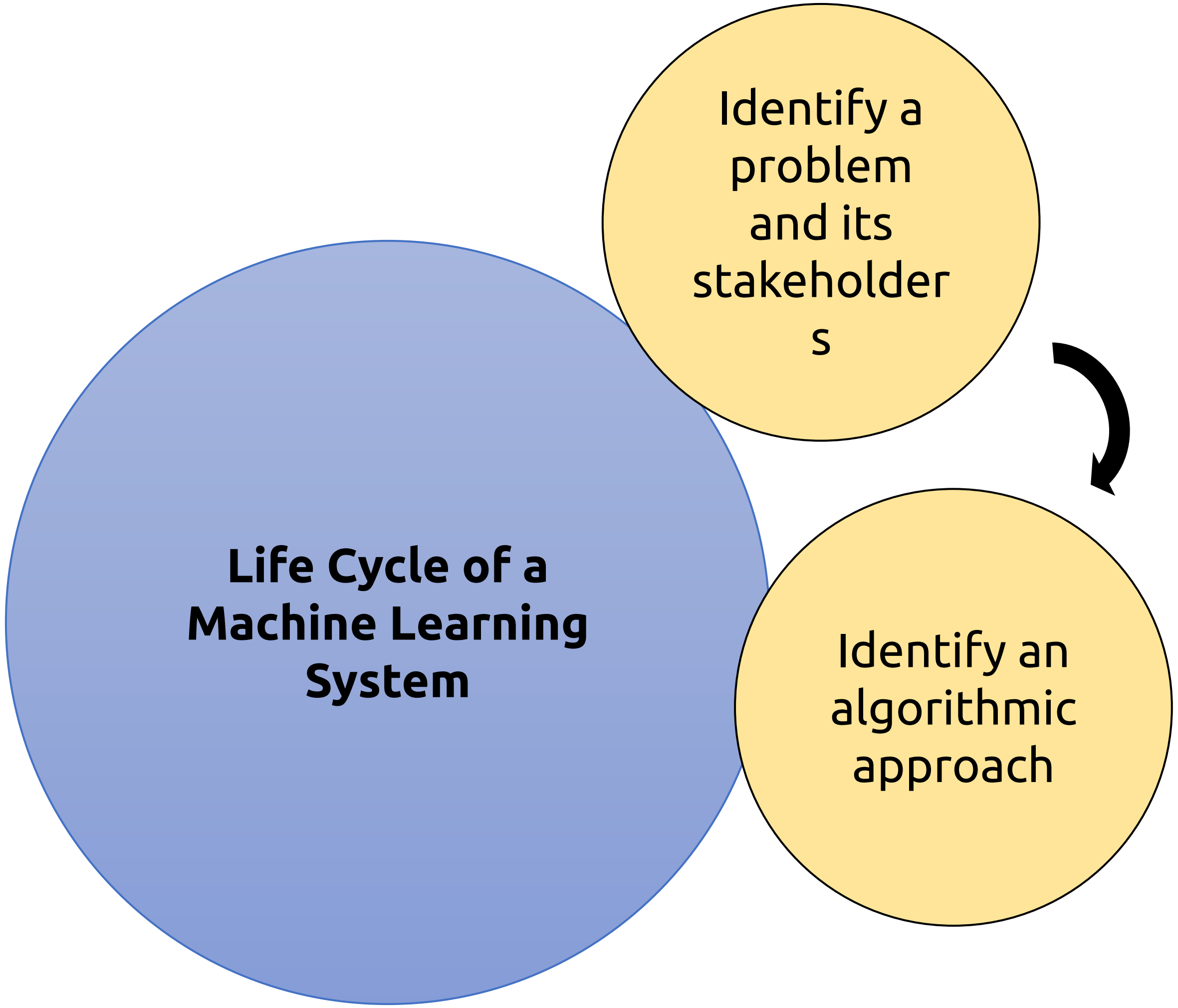
Example: digit recognition



- MNIST gives us tens of thousands of training examples
- Ok if predictions aren't explainable, as long as they're high accuracy
- Seems like a clear **yes**

Does deep learning make sense?

- Can we acquire enough training data?
- Are we ok with not understanding why the model makes the decisions it does?

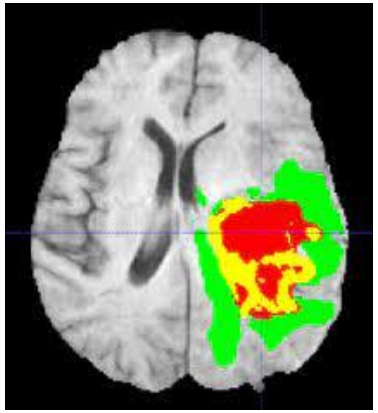


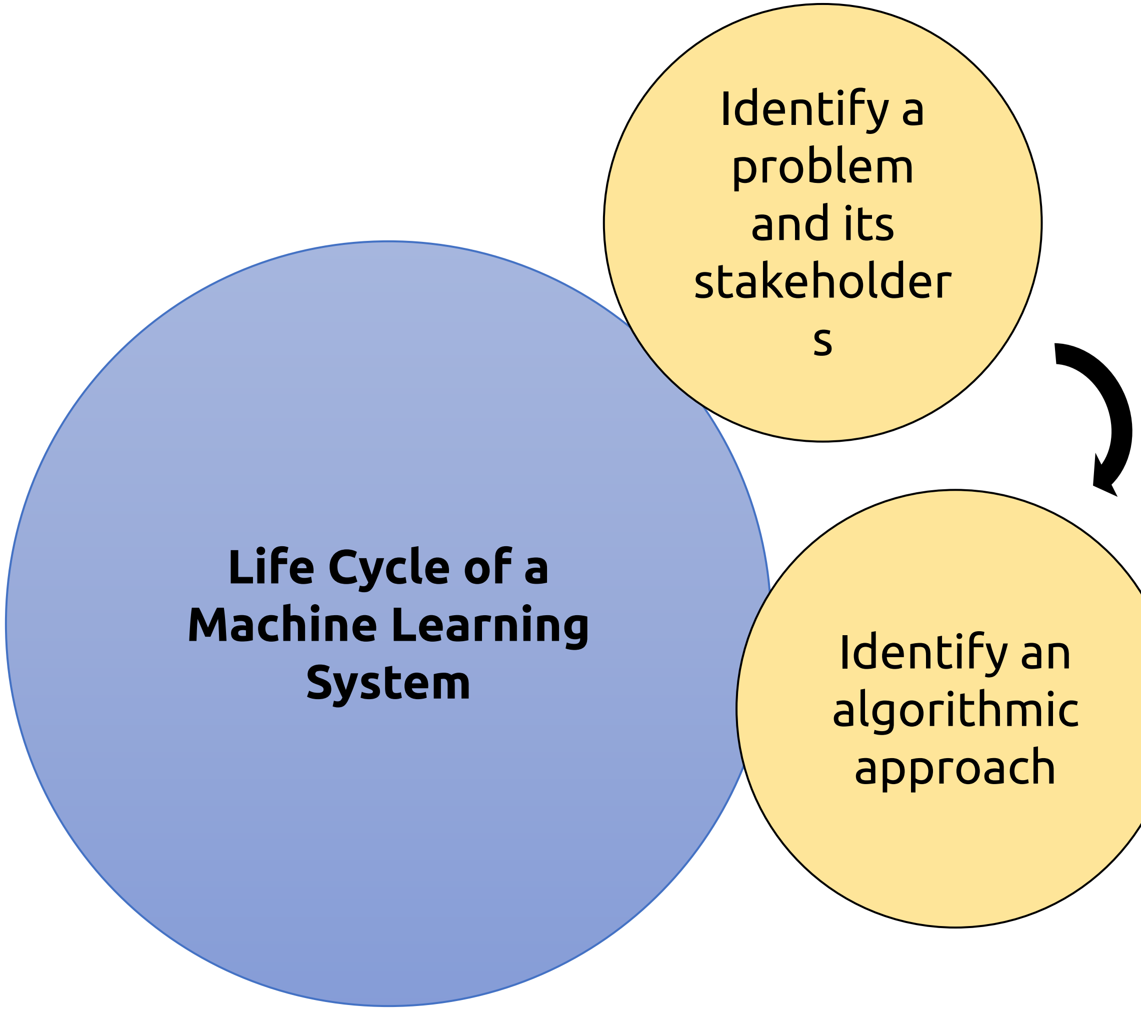
Example: personal health recommendations

- Can only gather a few datapoints from a single person's life
- DL might be a **no**; other, less data-intensive ML approaches might be viable

Does deep learning make sense?
- Can we acquire enough training data?
- Are we ok with not understanding why the model makes the decisions it does?

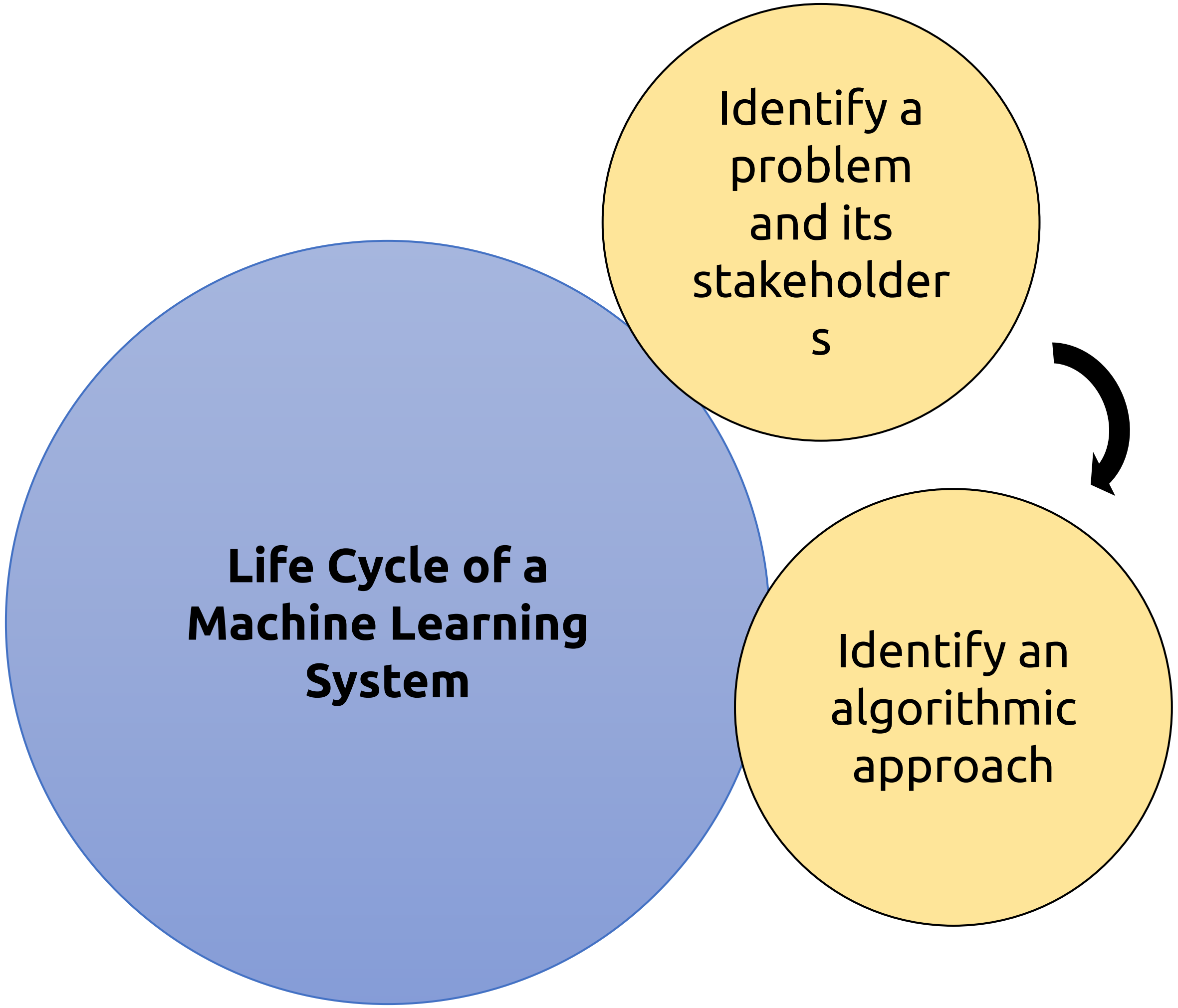
Example: Automatic brain tumor segmentation?





Does deep learning make sense?

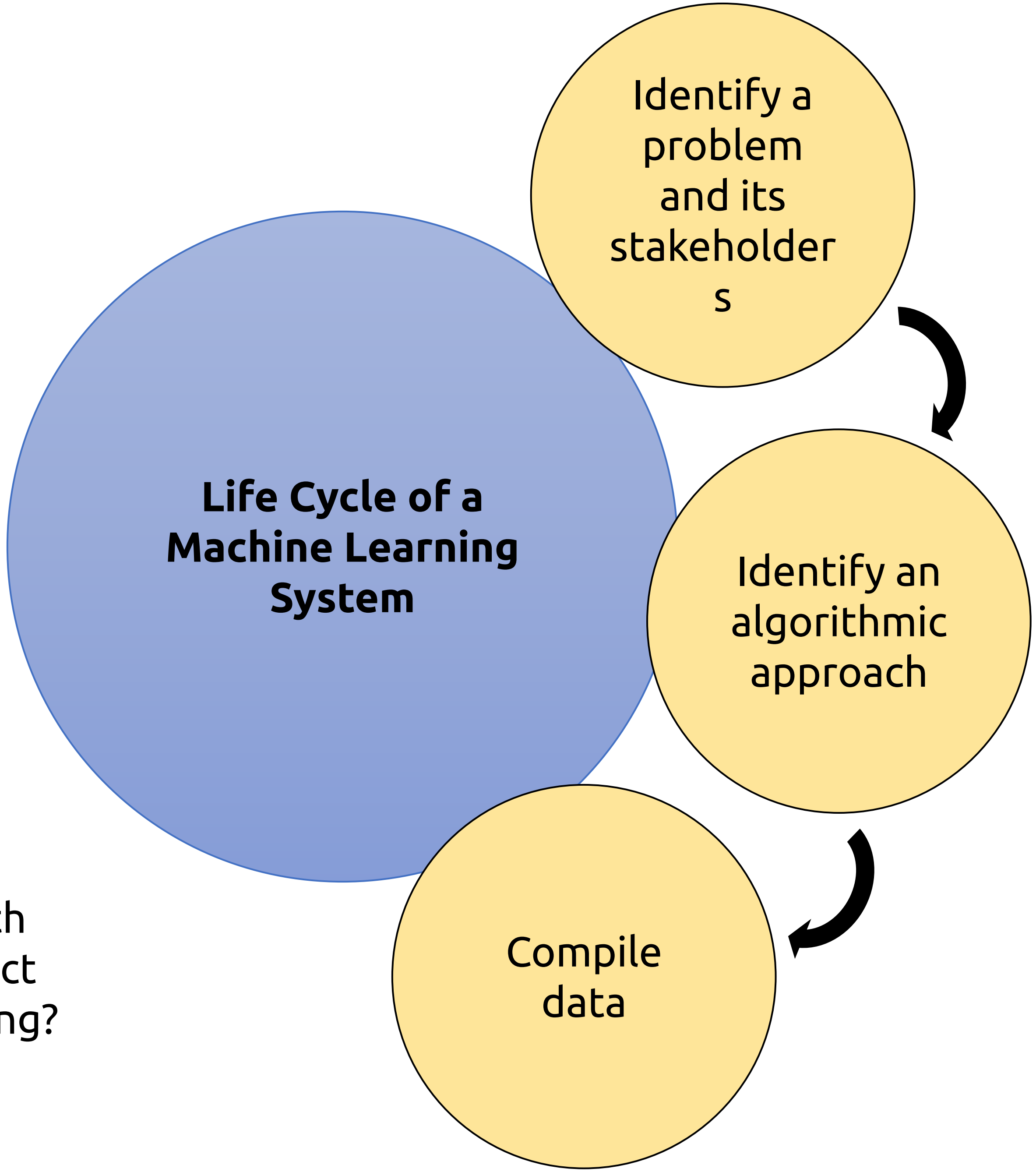
- Can we acquire enough training data?
- Are we ok with not understanding why the model makes the decisions it does?



Example: bail determination

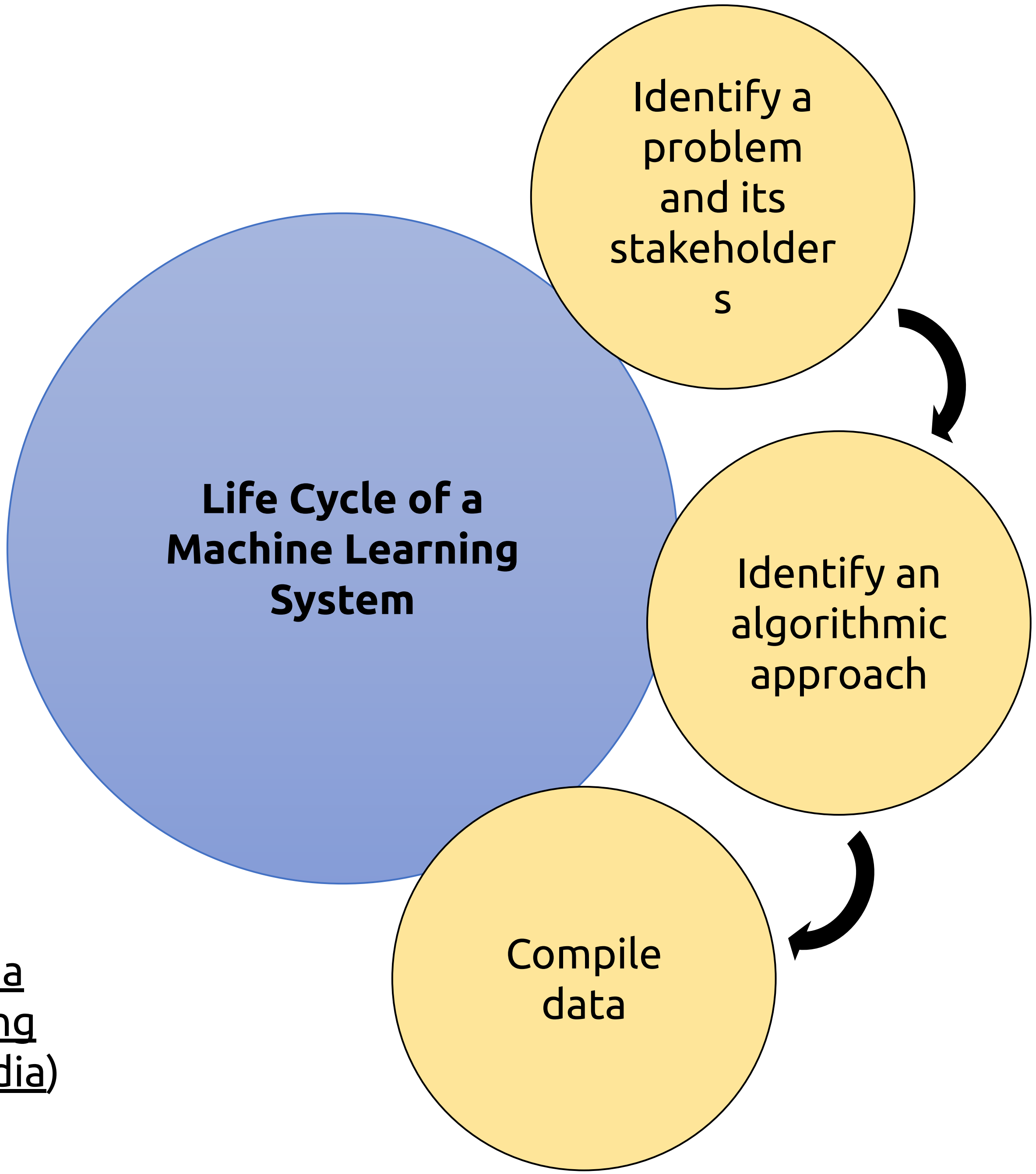
- If a machine denies someone bail, should it be required to explain why?
- DL might be a **no**; other, more interpretable ML models might

Does deep learning make sense?
- Can we acquire enough training data?
- Are we ok with not understanding why the model makes the decisions it does?



Is our data representative?

- Does it contain feature values with the same frequency that we expect to see those values when deploying?
- E.g. race, ethnicity, gender, …



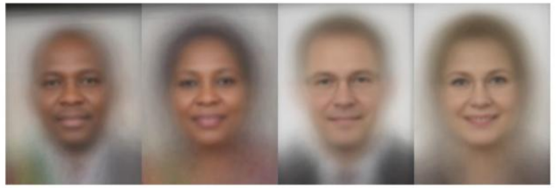
Was our data collected ethically?

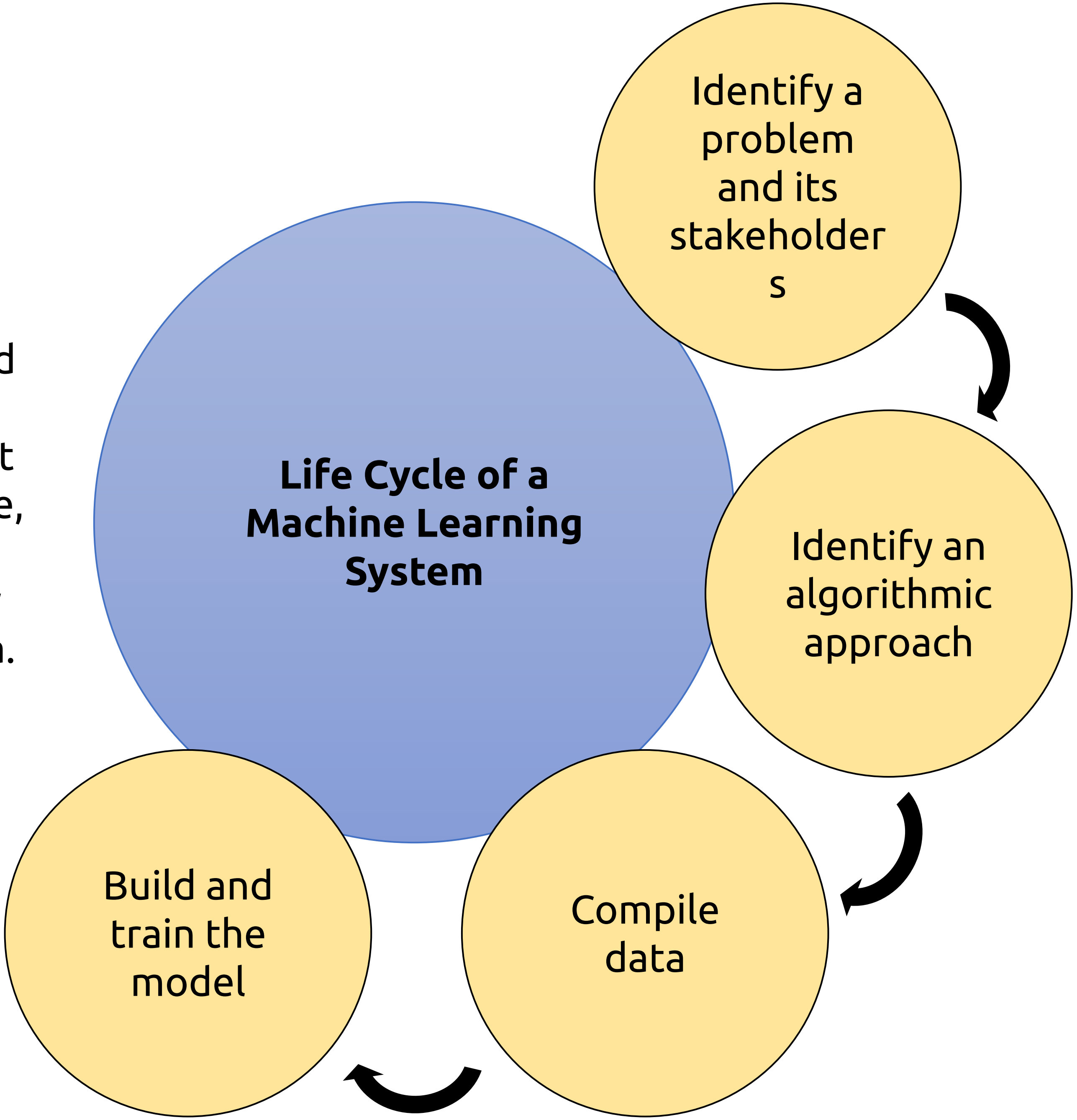
- Did we obtain consent, where appropriate?
- (E.g. you probably shouldn't build a face recognition service by scraping millions of photos from social media)

What's the objective/loss function?

- Does it reflect all of our desired outcomes and values?
- E.g. minimizing average test set error sounds good at face value, but can lead to systematic underperformance on minority subpopulations within the data.

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

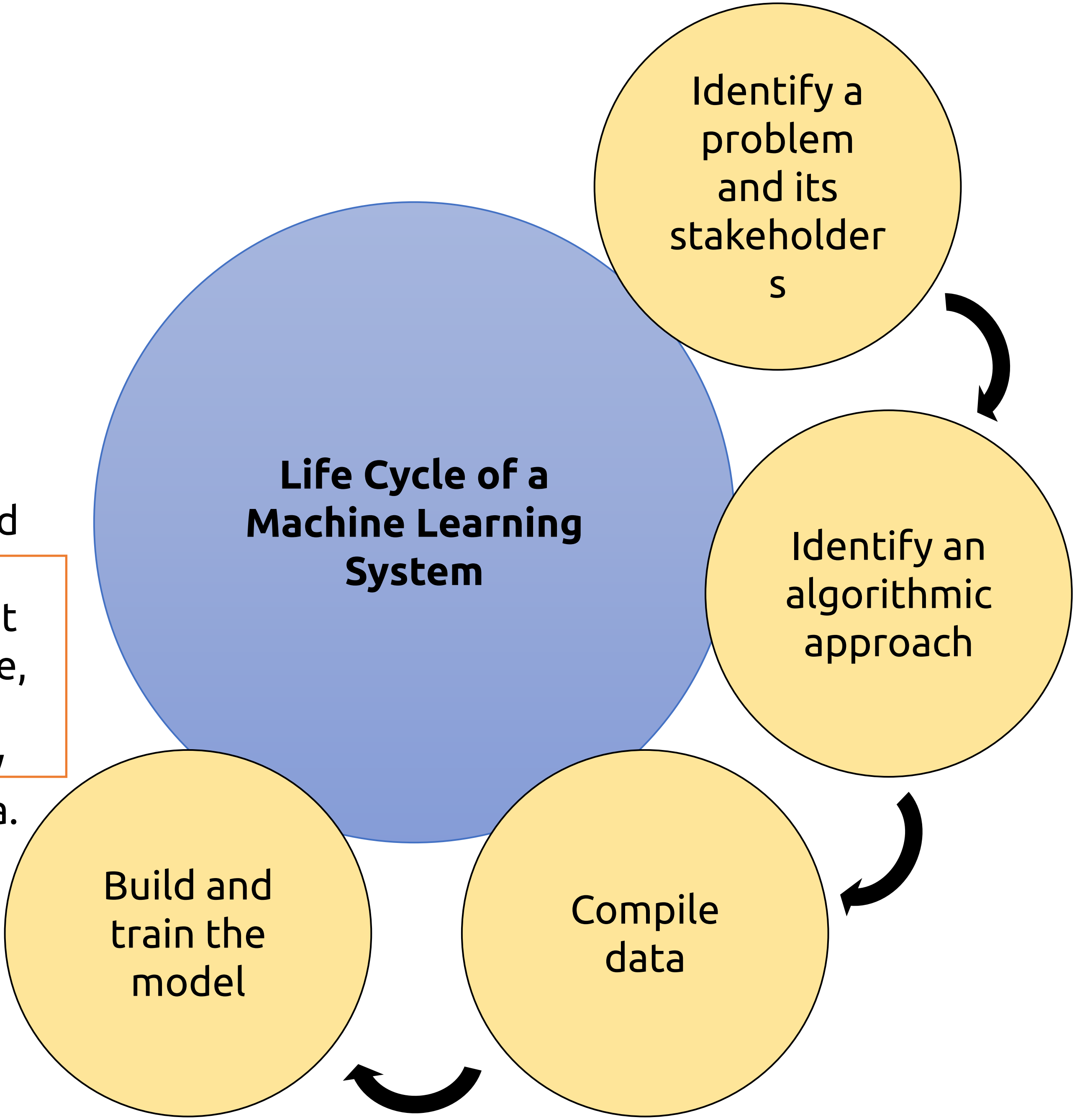




What's the objective/loss function?

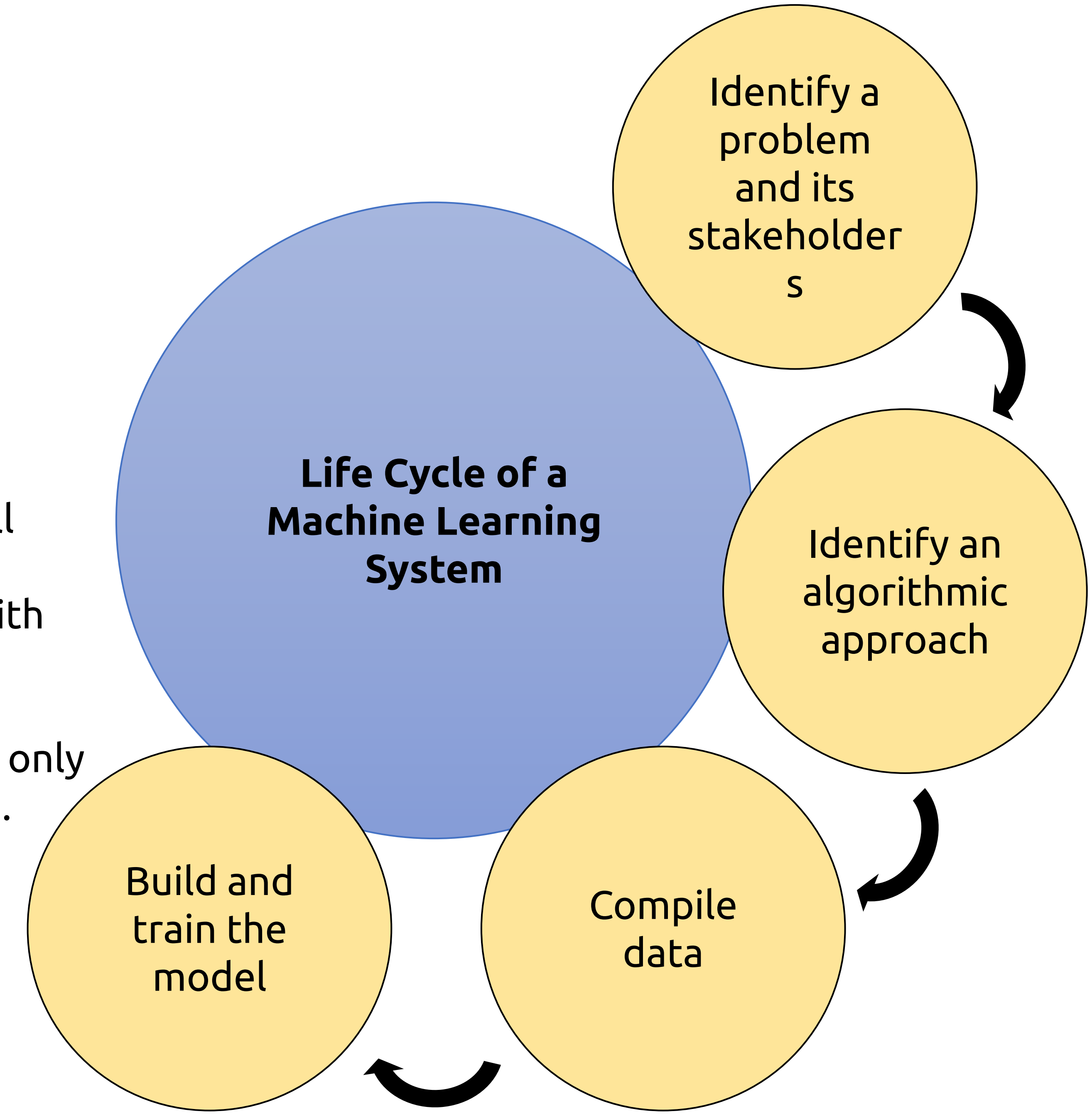
- Does it reflect all of our desired outcomes and values?
- E.g. minimizing average test set error sounds good at face value, but can lead to systematic underperformance on minority subpopulations within the data.

Algorithmic fairness

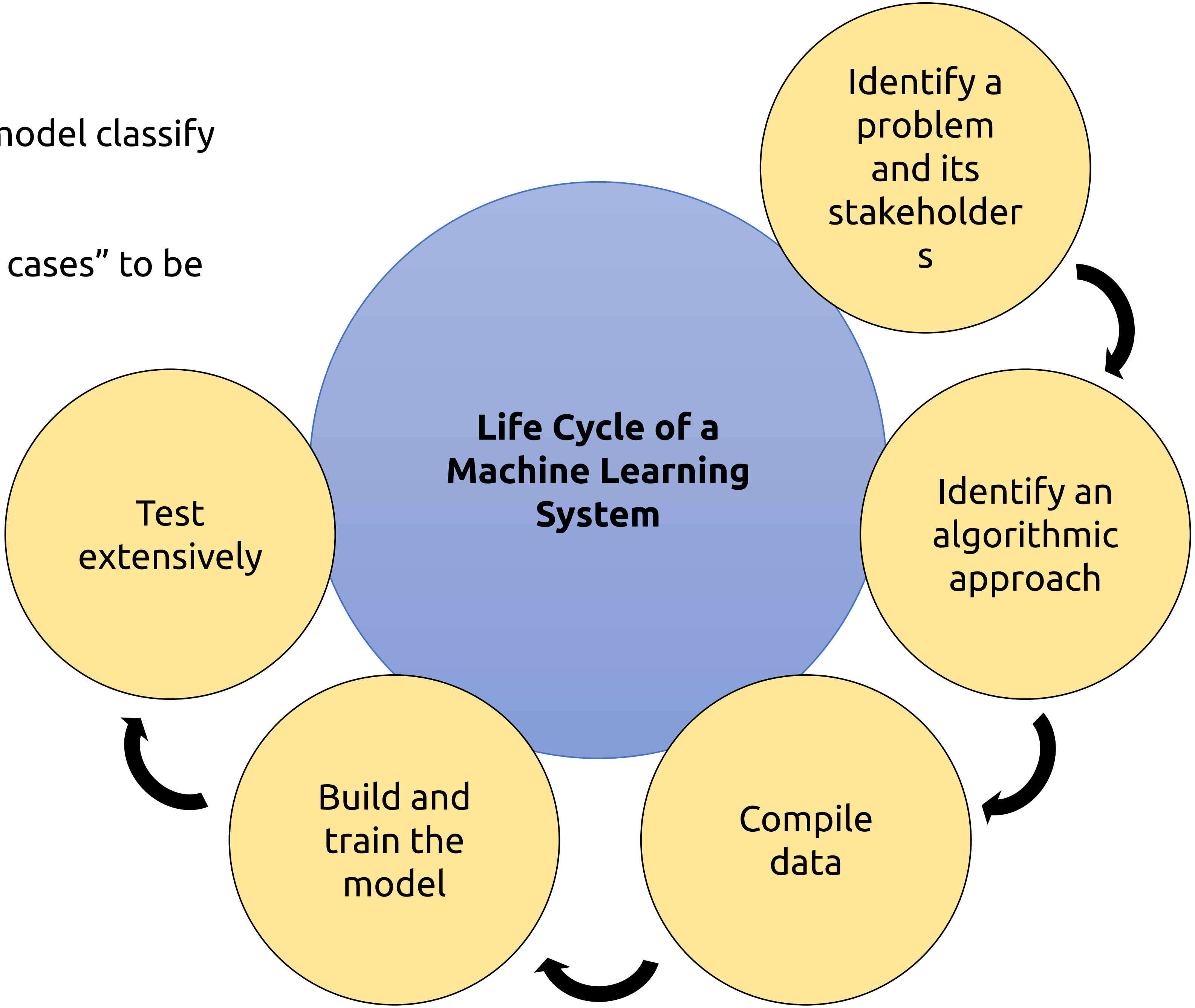


How do we label the training examples?

- Do these labels account for all possibilities?
- Are these labels consistent with the value system(s) of our stakeholders
- E.g. a “gender” attribute with only “male” and “female” options...



- What does the model classify correctly?
- Incorrectly?
- Are there “edge cases” to be aware of?

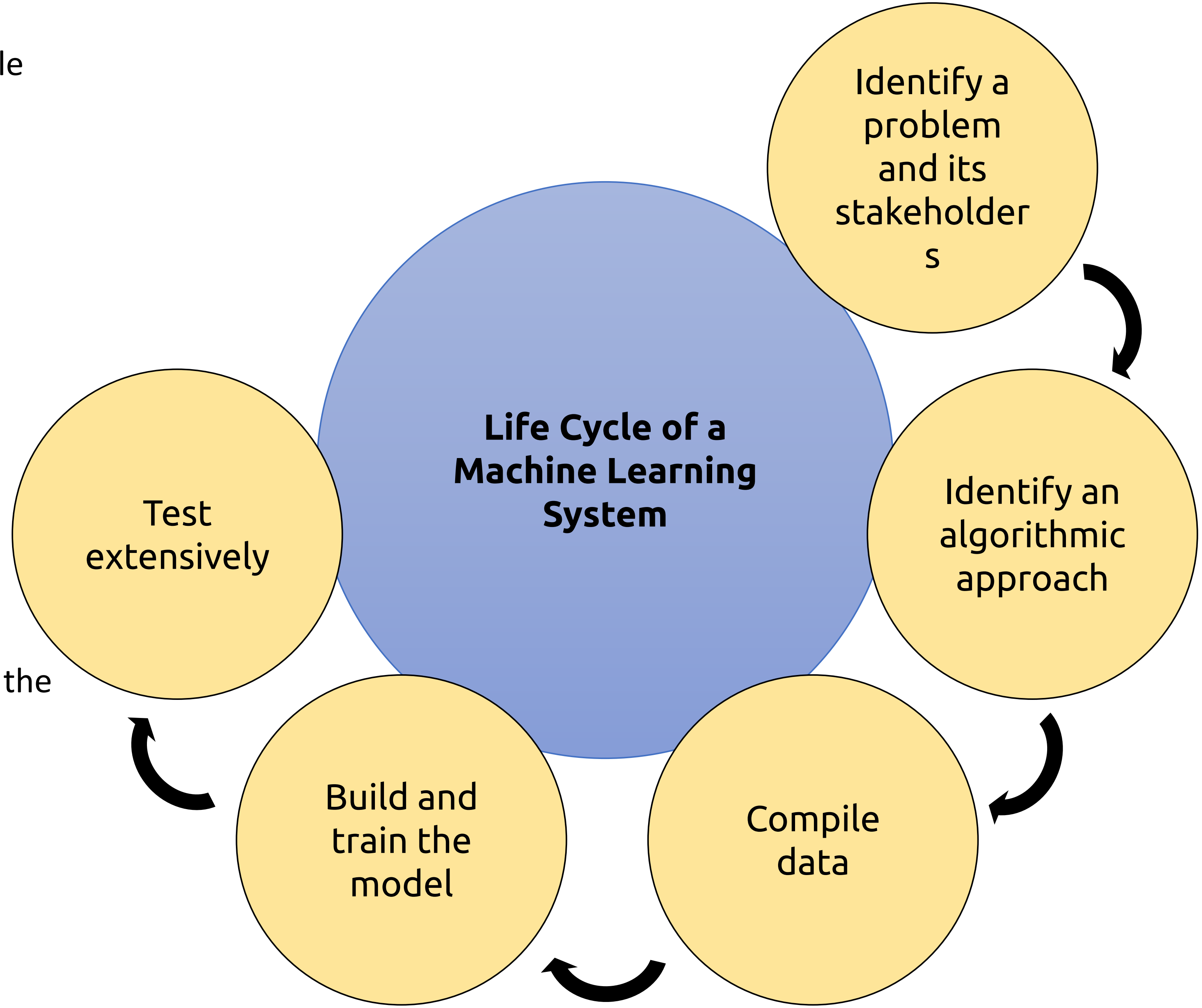


What is an acceptable margin of error?

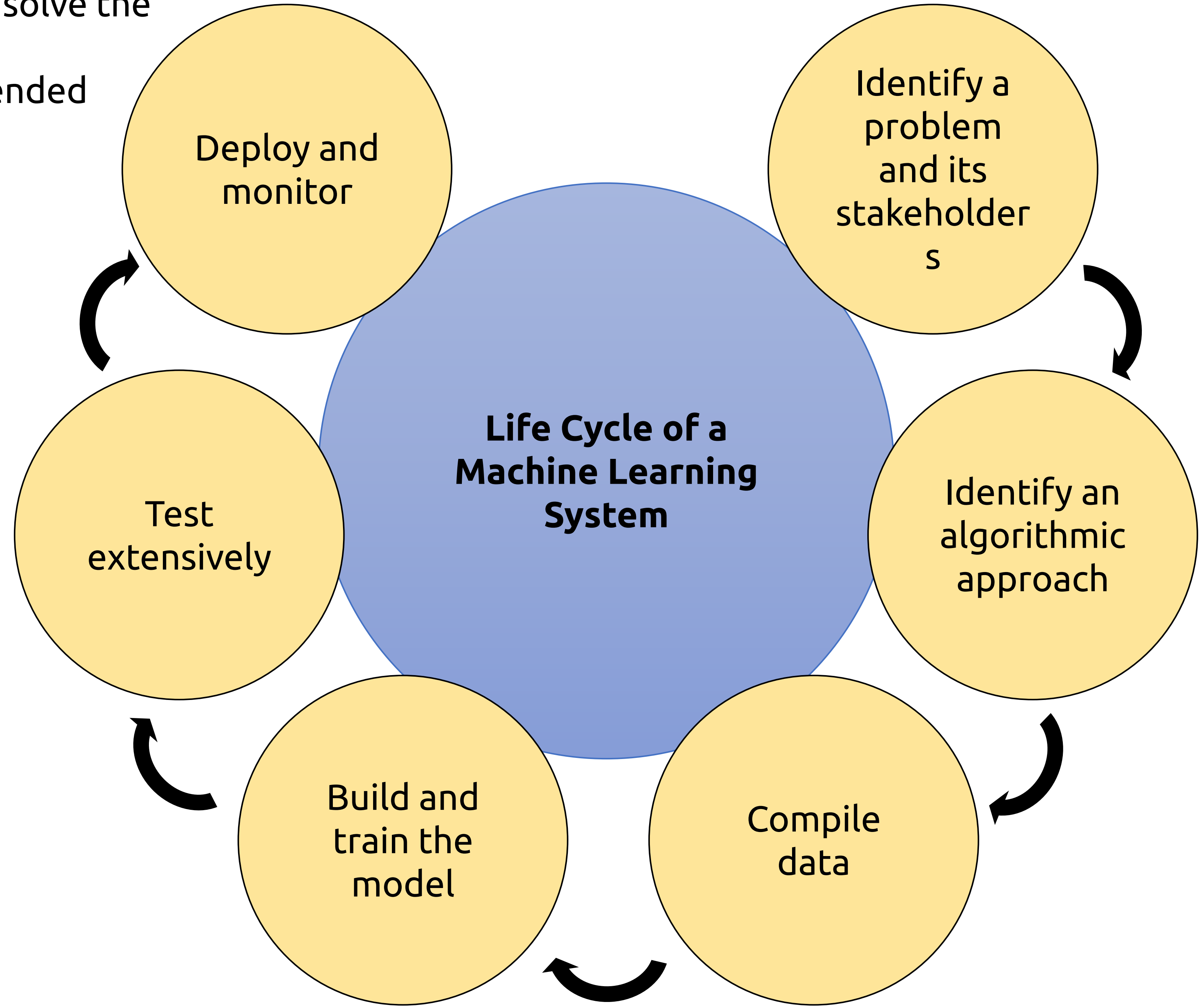
In different situations, incorrect predictions can have different consequences.

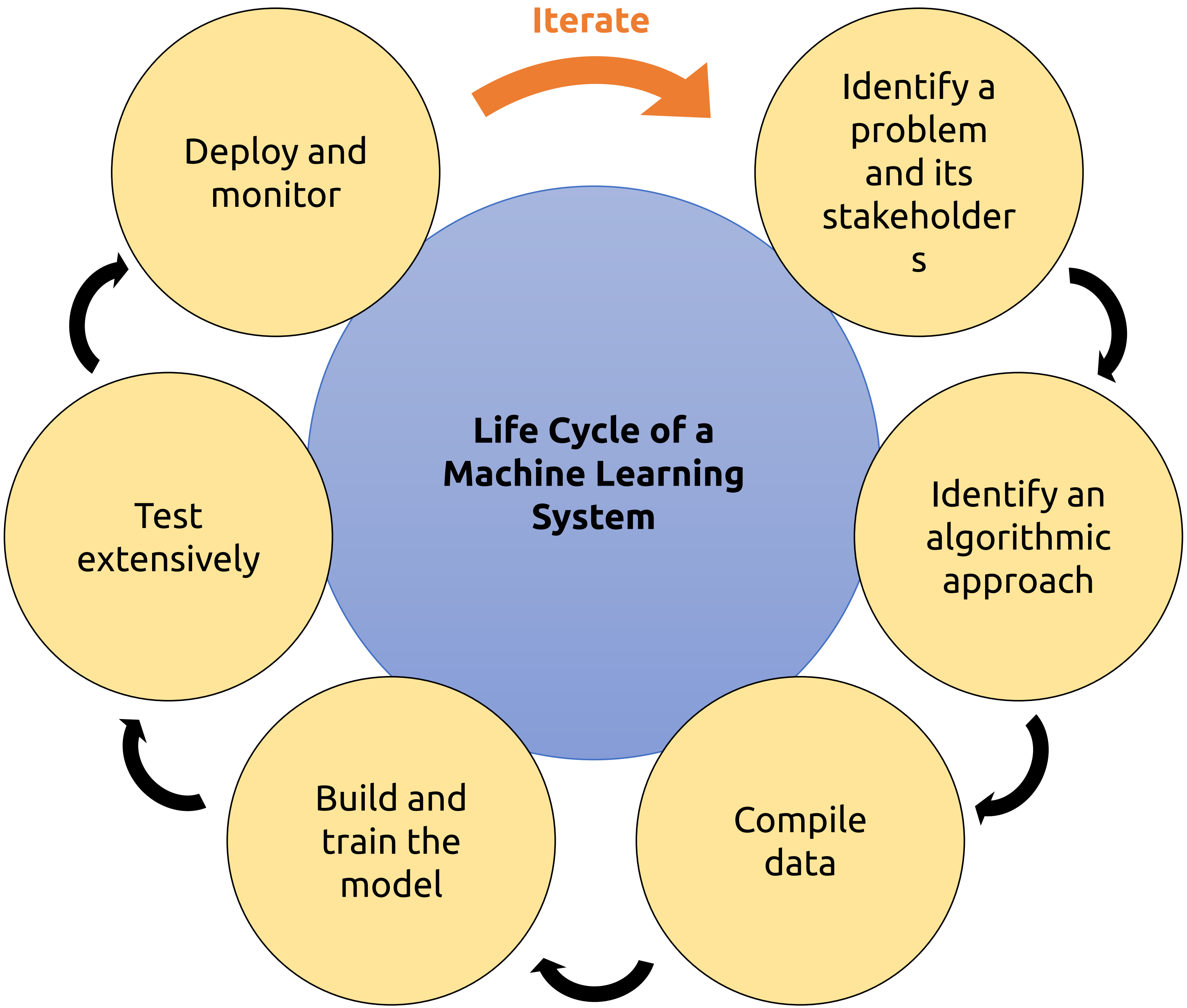
Incorrect prediction of a dog image for curation of images

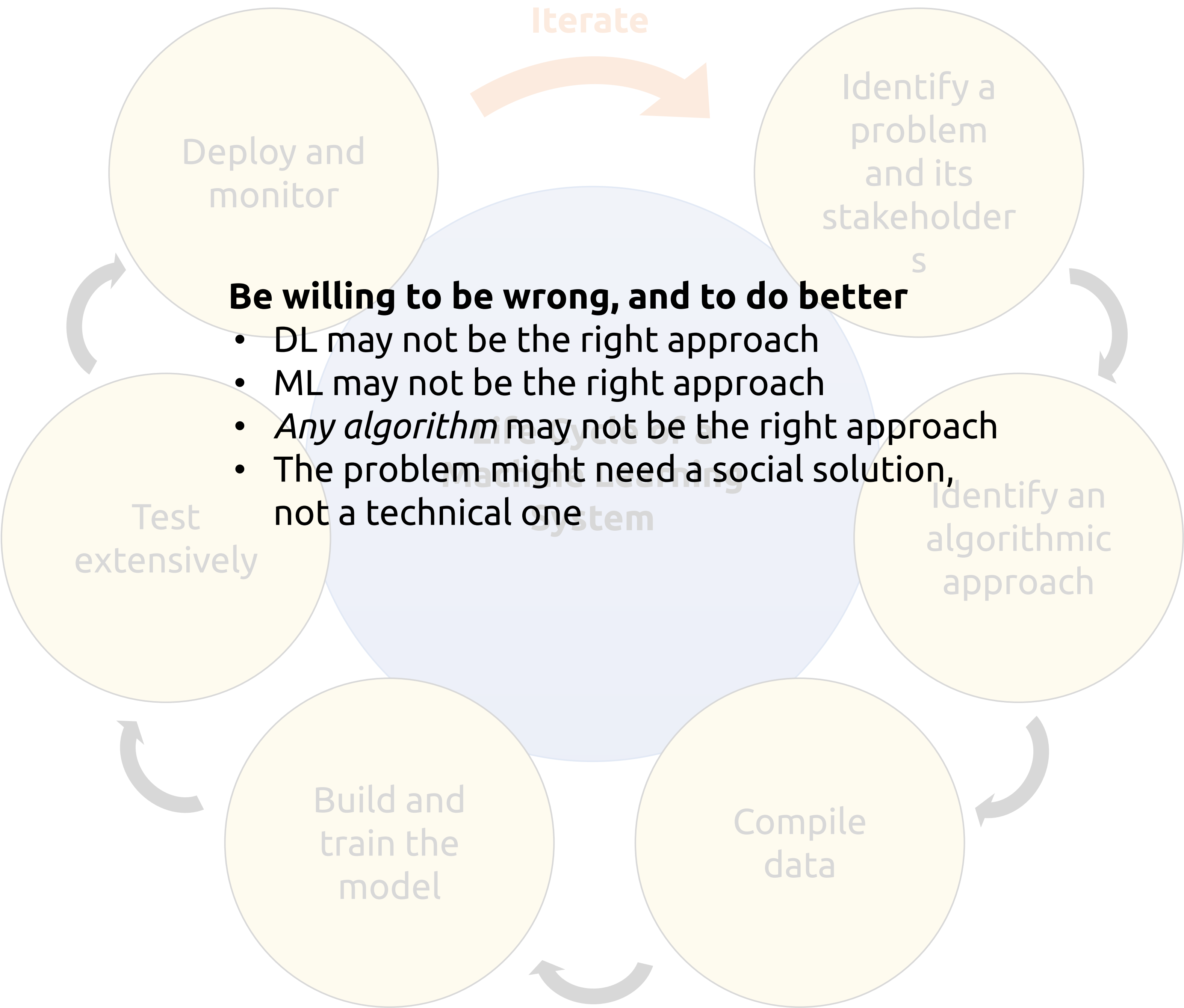
Incorrect prediction of the speed sign digits in autonomous vehicles



- Does our model solve the problem?
- Are there unintended consequences?




Iterate
Identify a problem and its stakeholder s
Identify an algorithmic approach
Compile data
Build and train the model
Test extensively
Deploy and monitor
Life Cycle of a Machine Learning System



Be willing to be wrong, and to do better

- DL may not be the right approach
- ML may not be the right approach
- *Any algorithm* may not be the right approach
- The problem might need a social solution, not a technical one

Finally: You’re not alone

- ***You*** won’t always necessarily have the background to answer all of these questions
- **Include domain/subject matter experts** who may be able to point out unforeseen consequences