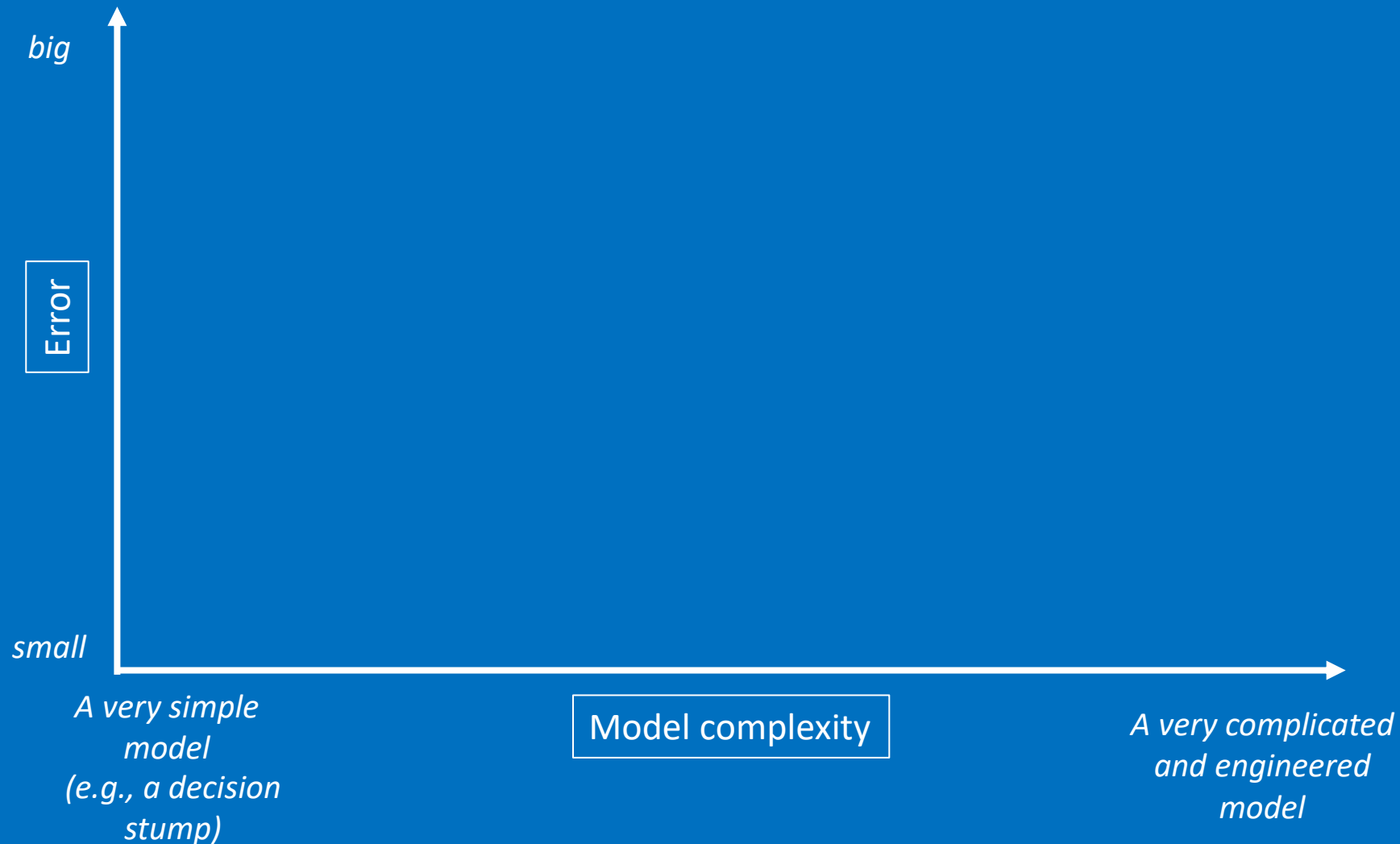# Overfitting and underfitting
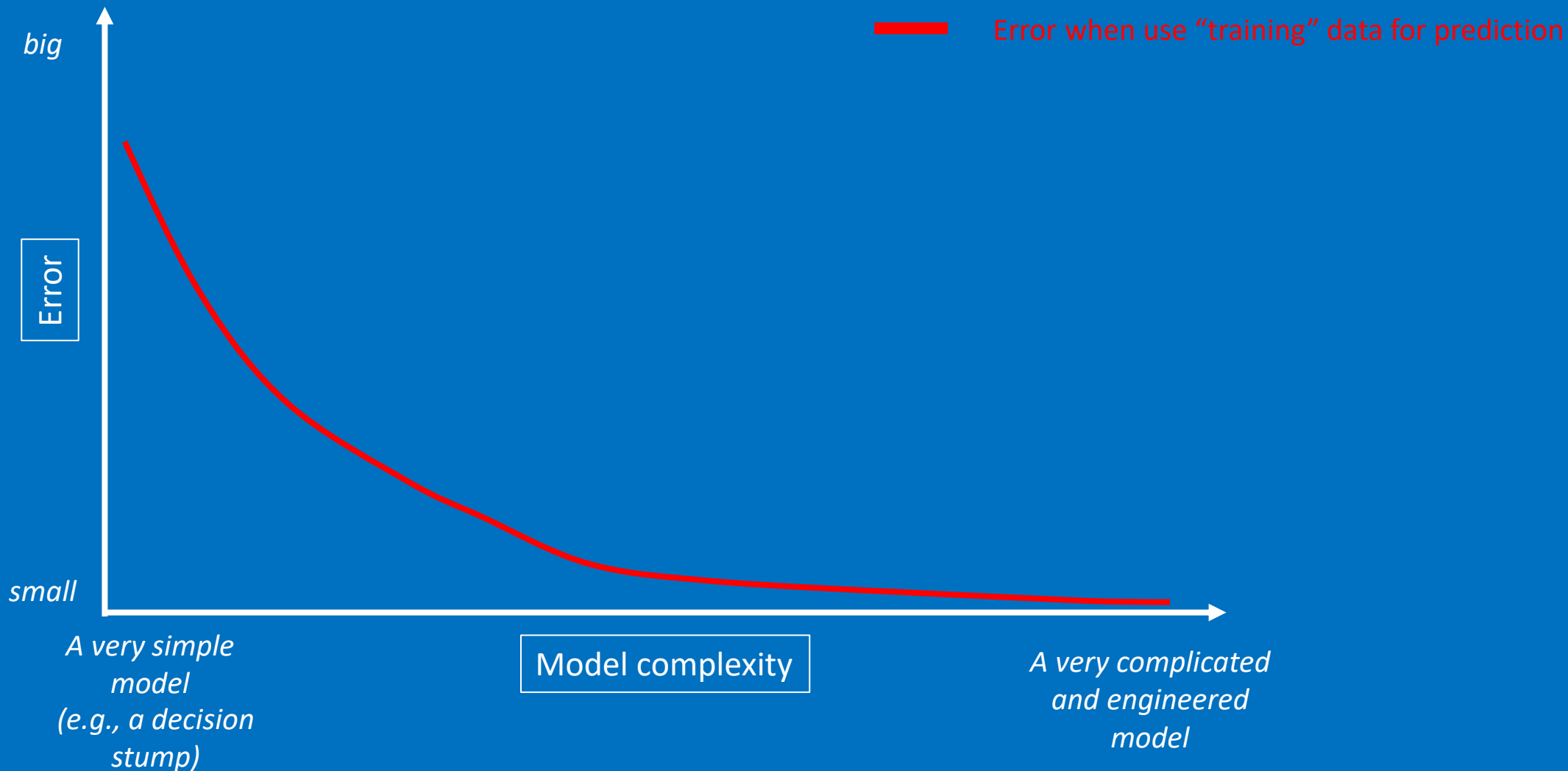
Sijin Zhang

The purpose of ML is to have a model which is capable of "predicting" well on "unseen" data

The purpose of ML is to have a model which is capable of "predicting" well on "unseen" data
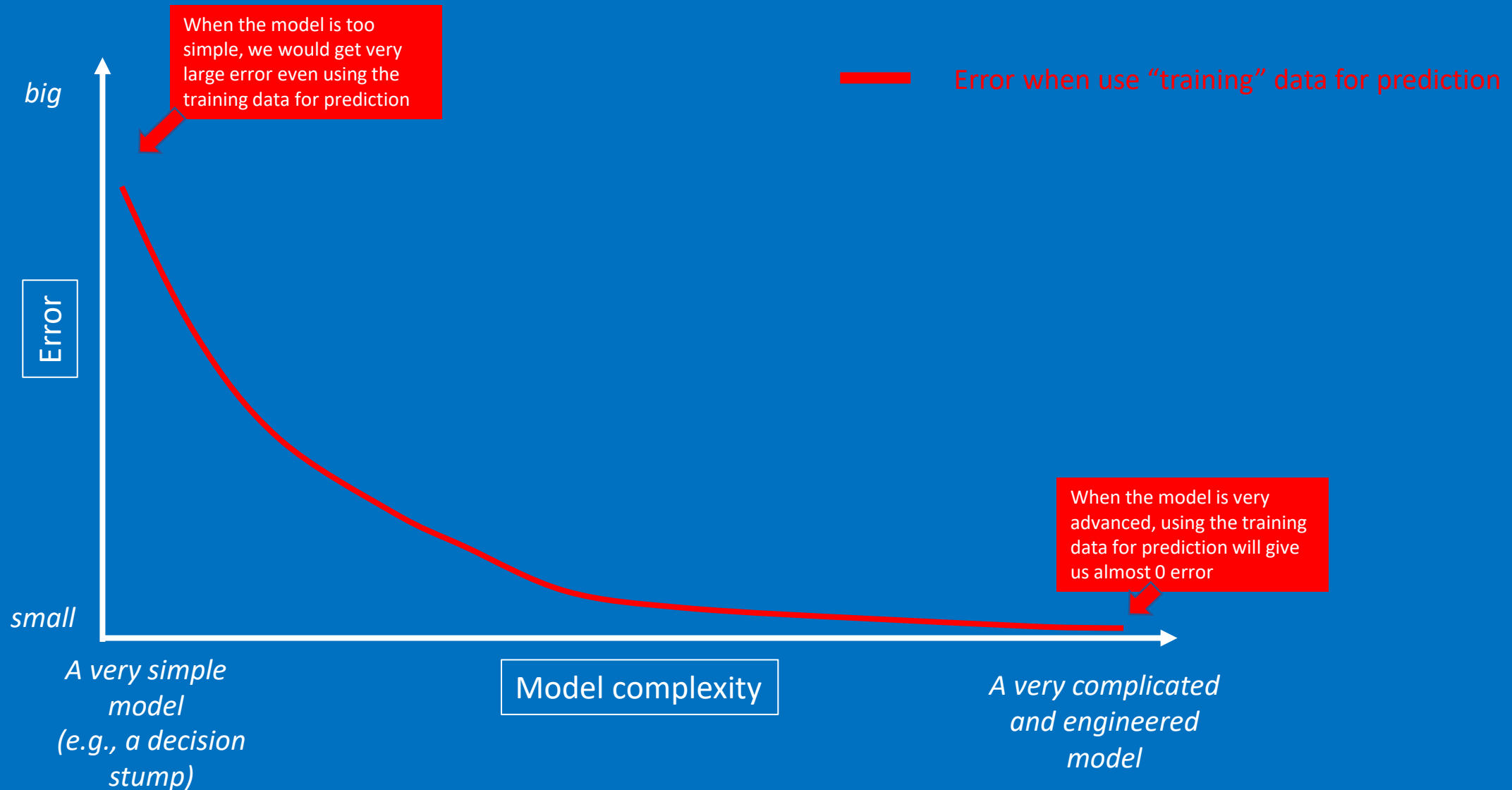
The purpose of ML is to have a model which is capable of "predicting" well on "unseen" data

The purpose of ML is to have a model which is capable of "<u>predicting</u>" well on "<u>unseen</u>" data

When the model is too simple, we would get very large error even using the training data for prediction

Error when use "training" data for prediction

big

Error

When the model is very advanced, using the training data for prediction will give us almost 0 error

small

A very simple model
(e.g., a decision stump)

Model complexity

A very complicated and engineered model

The purpose of ML is to have a model which is capable of "predicting" well on "unseen" data

When the model is too simple, we would get very large error even using the training data for prediction

Error when use "training" data for prediction

Error when use "test" data for prediction

big

Error

small

A very simple model (e.g., a decision stump)

Model complexity

A very complicated and engineered model

When the model is very advanced, using the training data for prediction will give us almost 0 error

The purpose of ML is to have a model which is capable of "predicting" well on "unseen" data

When the model is too simple, we would get very large error even using the training data for prediction

When the model is too simple, very large error in prediction with test dataset

━━━ Error when use "training" data for prediction

━━━ Error when use "test" data for prediction

big

Error

small

When the model is very advanced, using the training data for prediction will give us almost 0 error

A very simple model
(e.g., a decision stump)

Model complexity
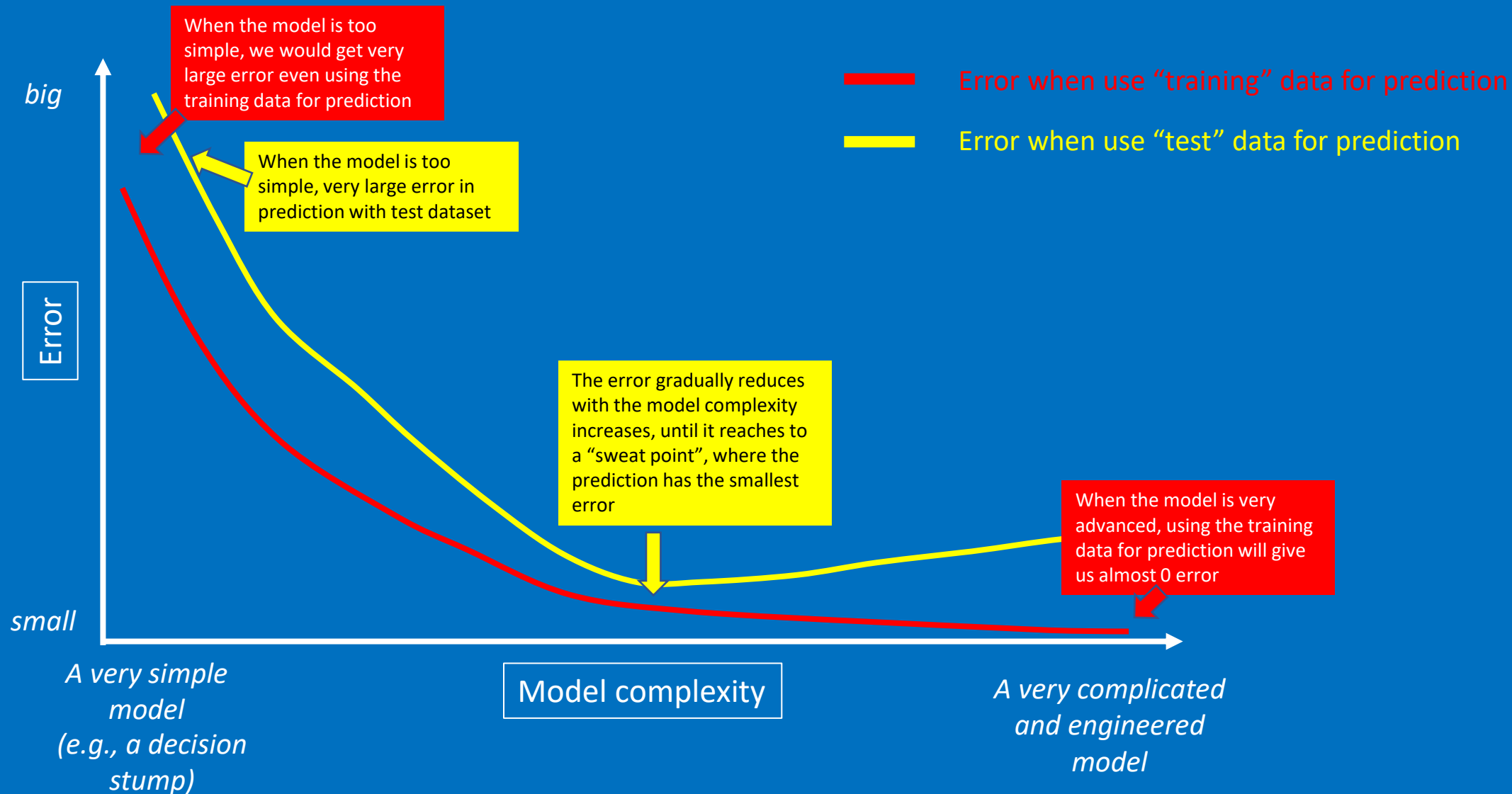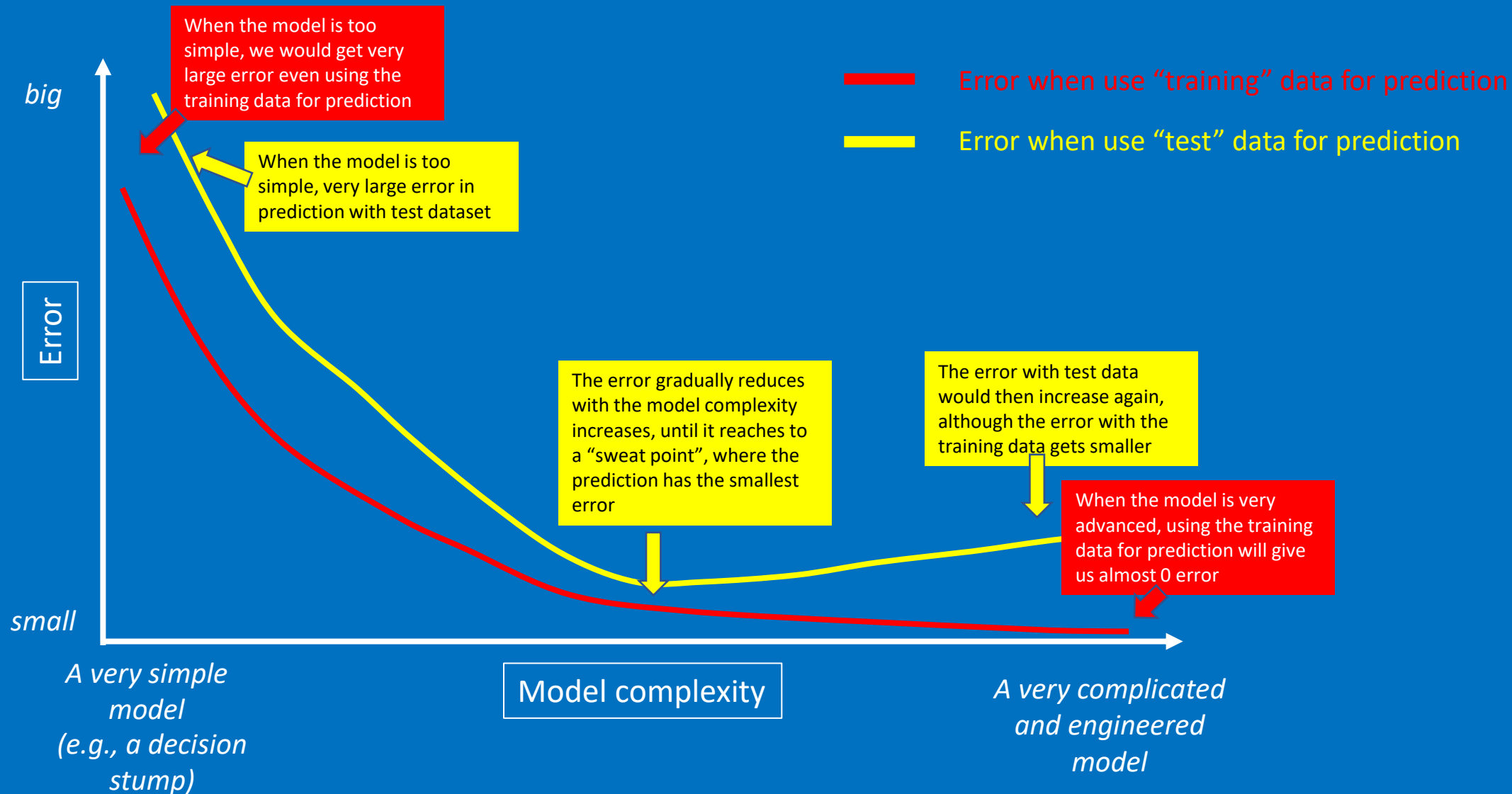
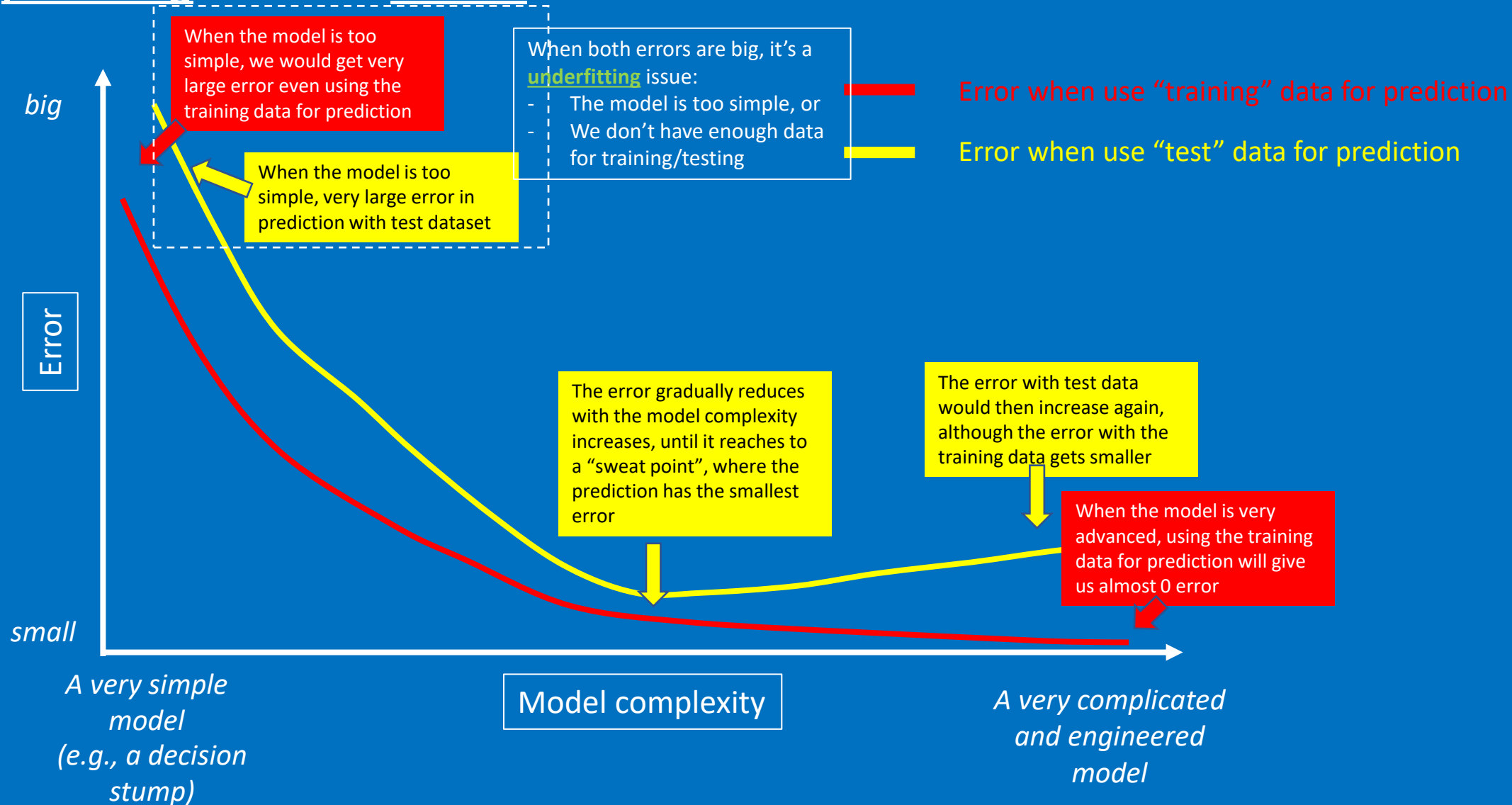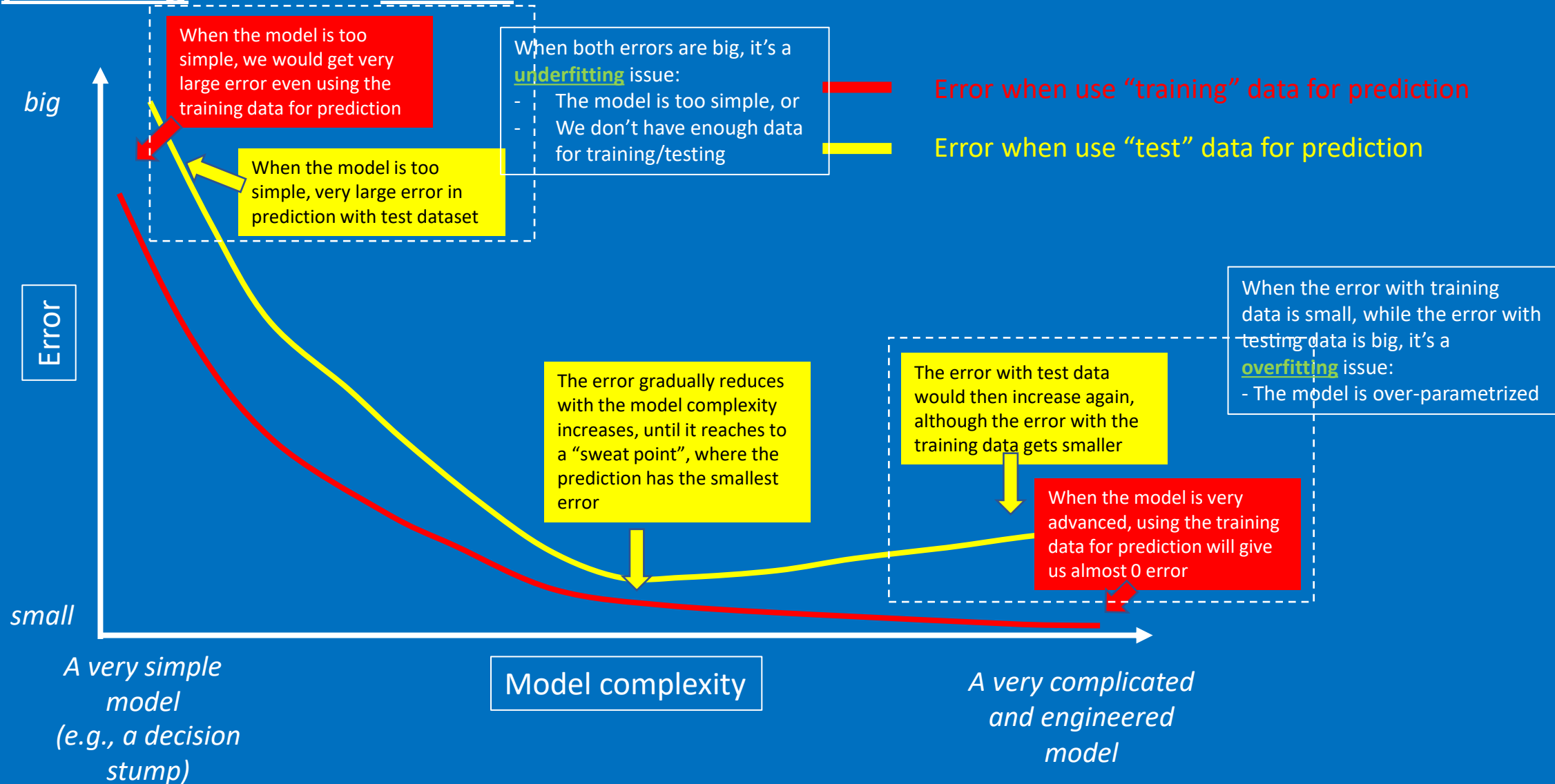A very complicated and engineered model

The purpose of ML is to have a model which is capable of "predicting" well on "unseen" data

The purpose of ML is to have a model which is capable of "predicting" well on "unseen" data

The purpose of ML is to have a model which is capable of "predicting" well on "unseen" data

big

Error

small

When the model is too simple, we would get very large error even using the training data for prediction

When both errors are big, it's a **underfitting** issue:
- The model is too simple, or
- We don't have enough data for training/testing

── Error when use "training" data for prediction

── Error when use "test" data for prediction

When the model is too simple, very large error in prediction with test dataset

The error gradually reduces with the model complexity increases, until it reaches to a "sweat point", where the prediction has the smallest error

The error with test data would then increase again, although the error with the training data gets smaller

When the model is very advanced, using the training data for prediction will give us almost 0 error

A very simple model (e.g., a decision stump)

Model complexity

A very complicated and engineered model

The purpose of ML is to have a model which is capable of "predicting" well on "unseen" data