

# Explainable AI

Cheat sheet [ex.pegg.io](https://ex.pegg.io) v.0.2

Start

Explaining a model's decisions?

No

Ah, so we're exploring the model and its internals

Yes

Is the model interpretable by design?

No

So less-interpretable methods perform better for your use-case? We understand!

Yes

Good choice! Especially for high-stakes decisions

## Neural Representations

SVCCA

Probes

Activation maximization

Feature visualization

TCAV

## Example Based

Adversarial

Influence functions

Counterfactual

## Model Specific

Gradient saliency

Integrated gradients

Attention

## Model Agnostic

SHAP

Perturbation

LIME

## Interpretable Models

Logistic regression

KNN

Linear models

Need a method that works on all models?

Yes

No

Also yes

Using special examples from the dataset could uncover insights about the model