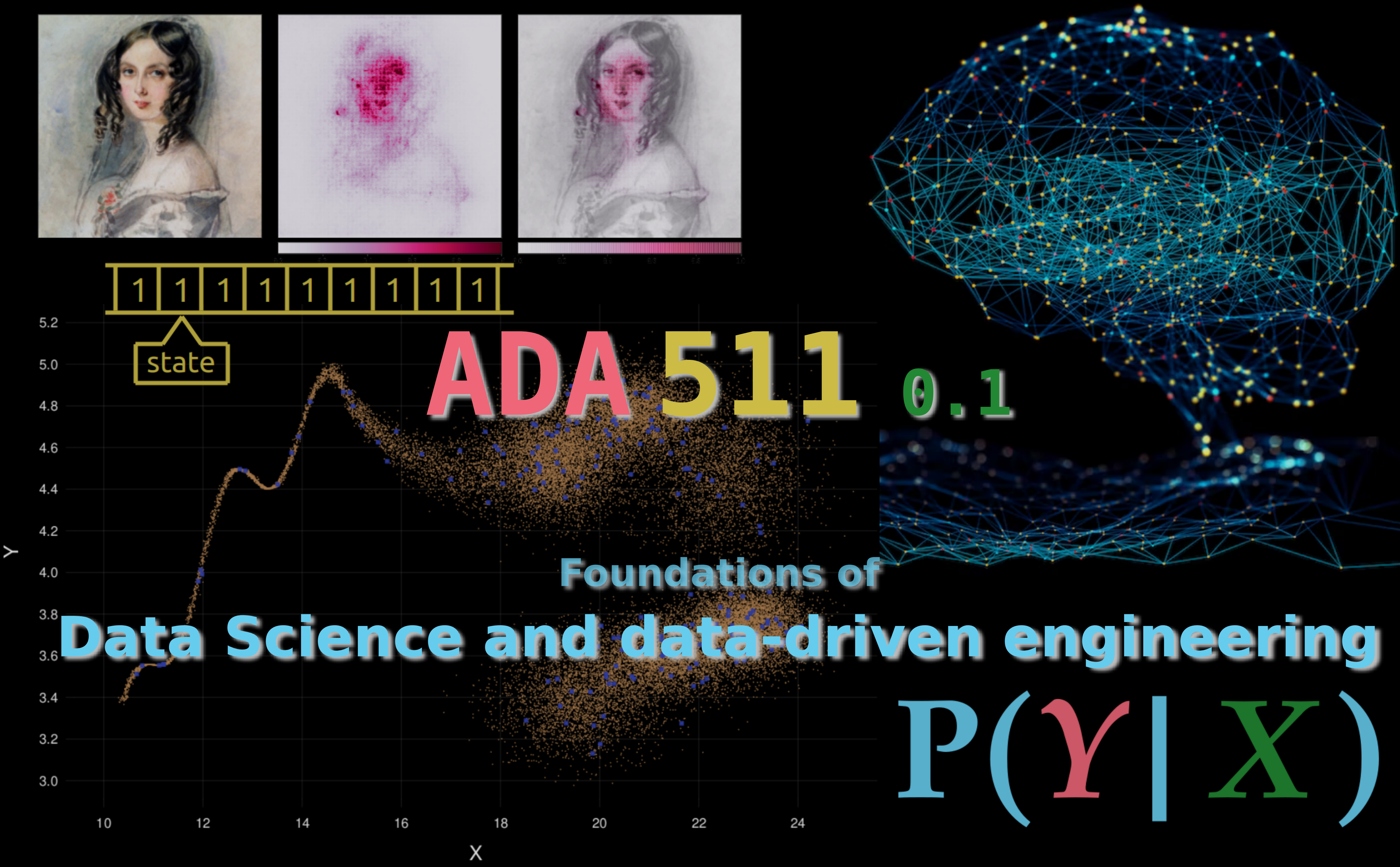
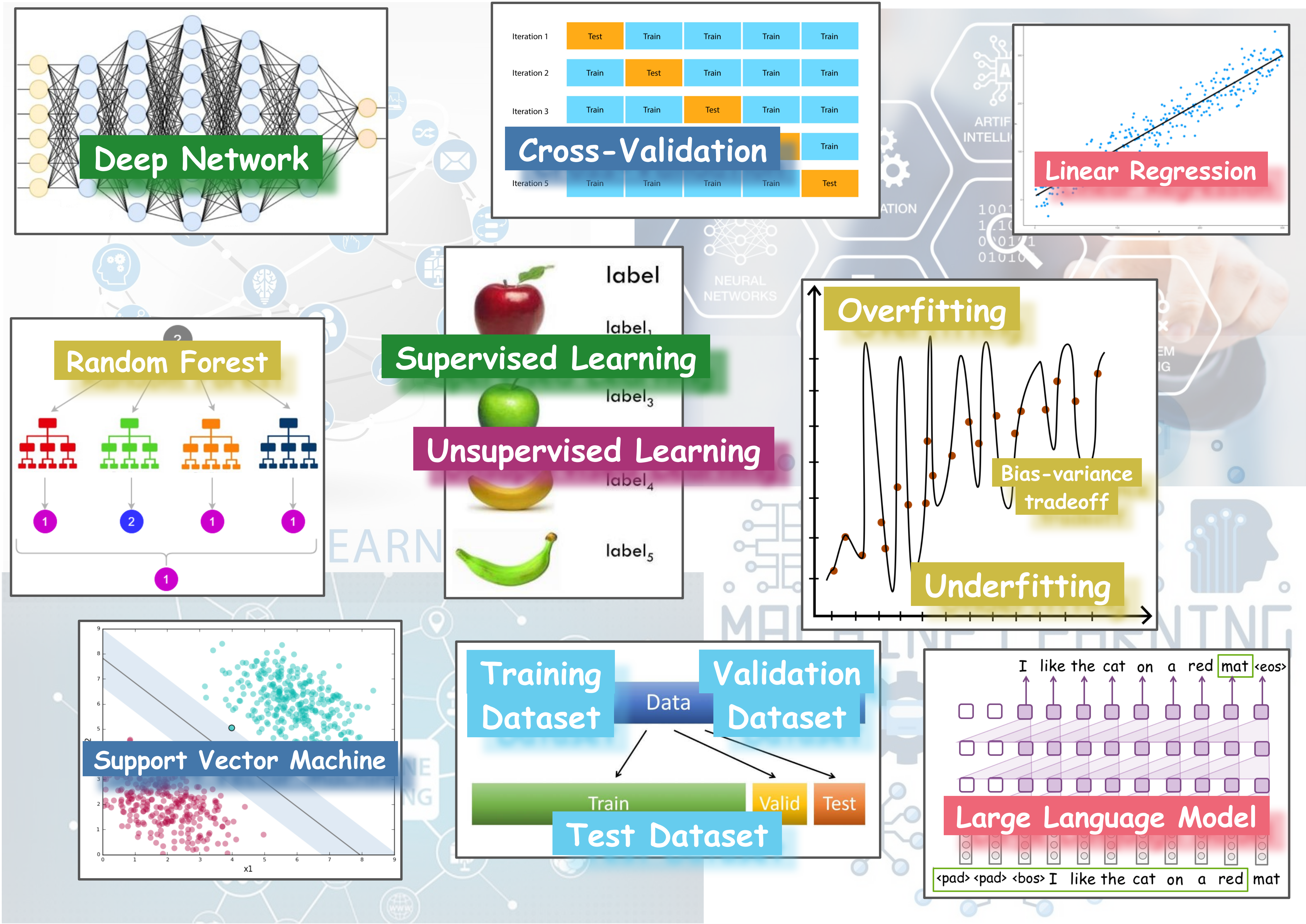


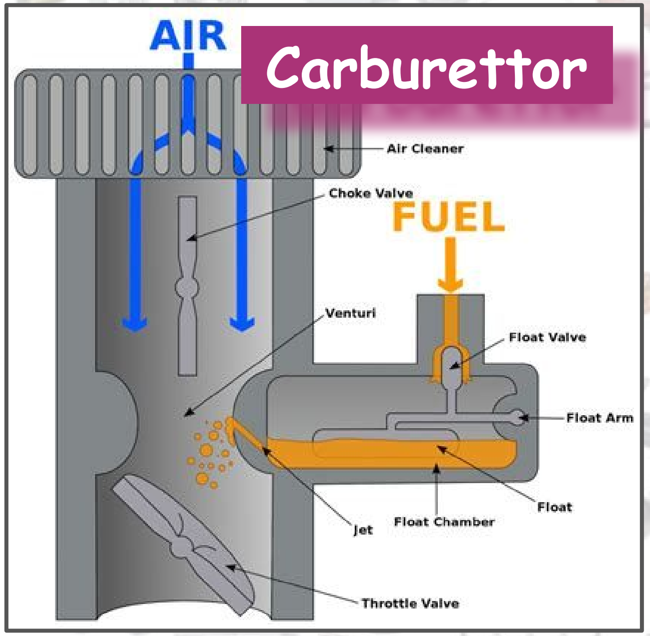
ADA 511 0.1

Foundations of Data Science and data-driven engineering

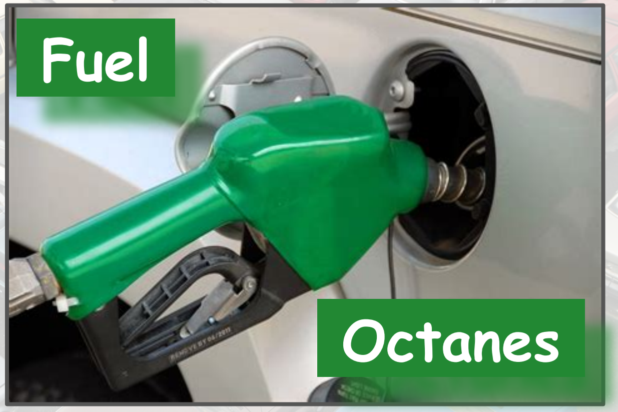
$$P(Y|X)$$


Deep Network
Cross-Validation
Iteration 1 Test Train Train Train Train
Iteration 2 Train Test Train Train Train
Iteration 3 Train Train Test Train Train
Train
Iteration 5 Train Train Train Train Test
Linear Regression
Random Forest
Supervised Learning
Unsupervised Learning
label
label$_1$
label$_3$
label$_4$
label$_5$
Overfitting
Bias-variance tradeoff
Underfitting
Training Dataset
Data
Validation Dataset
Train
Valid
Test
Test Dataset
Support Vector Machine
x1
I like the cat on a red mat <eos>
Large Language Model
<pad> <pad> <bos> I like the cat on a red mat

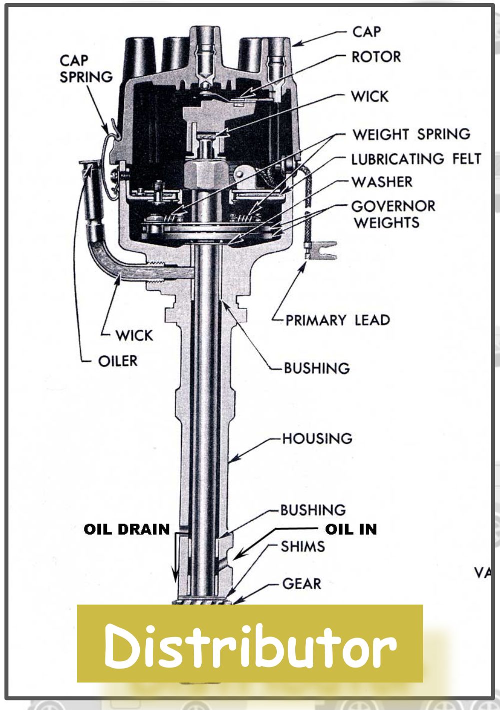
GTI
WOB-VS 99
$C_8H_{18} + 12.5\ O_8 \Rightarrow 8\ CO_2 + 9\ H_2O$
Isooctane
Oxygen
Carbon dioxide
Water
SEDAN
HATCHBACK
CUV
SUV
MICRO
CABRIOLET
SUPERCAR
COUPE
VAN
MINIVAN
CAMPERVAN
TRUCK
BIGTRUCK

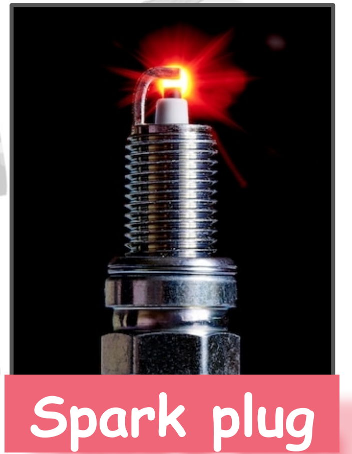

Carburettor
AIR
FUEL
Air Cleaner
Choke Valve
Venturi
Float Valve
Float Arm
Float
Float Chamber
Jet
Throttle Valve


WOB-VS 99
GTI
$C_8H_{18} + 12.5\ O_8 \Rightarrow 8\ CO_2 + 9\ H_2O$
Isooctane
Oxygen
Carbon dioxide
Water

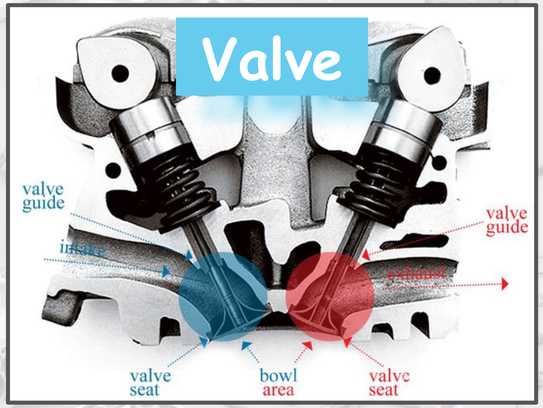

Fuel
Octanes

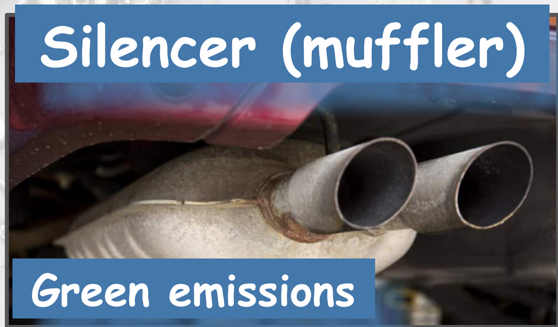

E
F
Fuel gauge

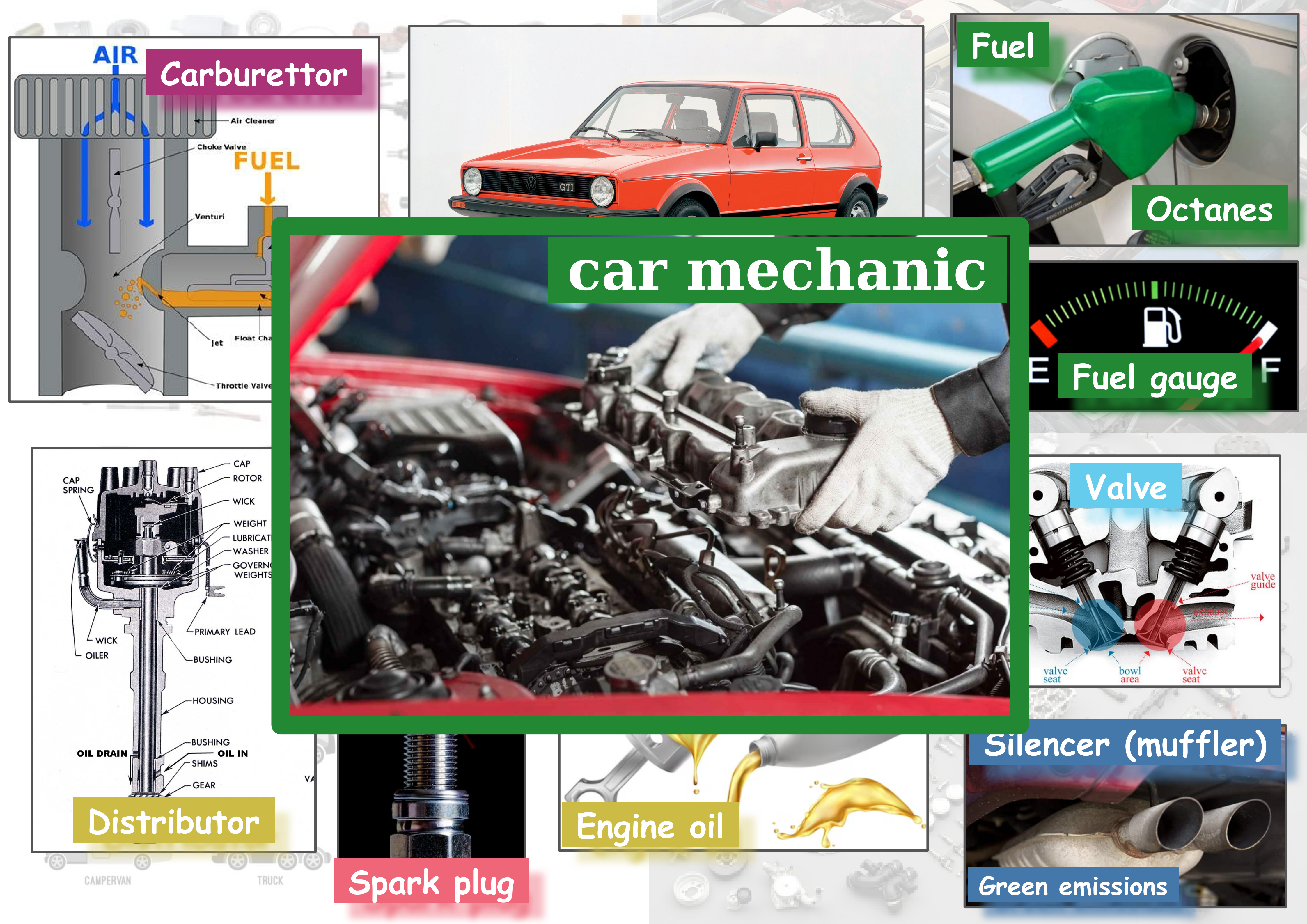

CAP
ROTOR
CAP SPRING
WICK
WEIGHT SPRING
LUBRICATING FELT
WASHER
GOVERNOR WEIGHTS
PRIMARY LEAD
WICK
OILER
BUSHING
HOUSING
BUSHING
OIL DRAIN
OIL IN
SHIMS
GEAR
Distributor


Spark plug


Engine oil


Valve
valve guide
valve guide
intake
exhaust
valve seat
bowl area
valve seat


Silencer (muffler)
Green emissions


car mechanic
Carburettor
AIR
Air Cleaner
Choke Valve
FUEL
Venturi
Jet
Throttle Valve
Fuel
Octanes
E
Fuel gauge
F
Valve
valve guide
valve seat
bowl area
valve seat
CAP
ROTOR
CAP SPRING
WICK
WEIGHT
WASHER
PRIMARY LEAD
WICK
OILER
BUSHING
HOUSING
BUSHING
OIL DRAIN
OIL IN
SHIMS
GEAR
Distributor
CAMPERVAN
TRUCK
Spark plug
Engine oil
Silencer (muffler)
Green emissions

cartwright

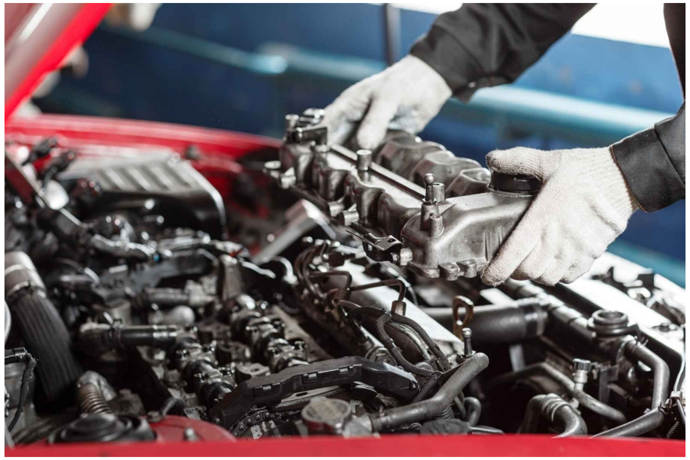
car mechanic

EV mechanic

?

automotive engineer

car mechanic



automotive engineer

specific

practical

procedural

technology-dependent

...

knowledge

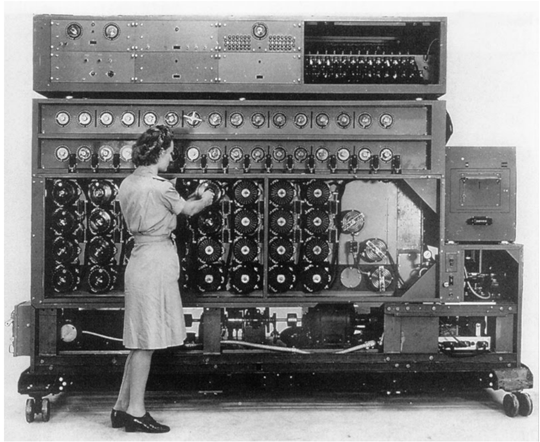
general

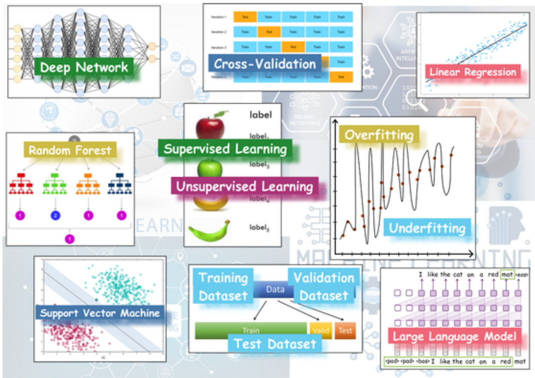
principled

exploratory

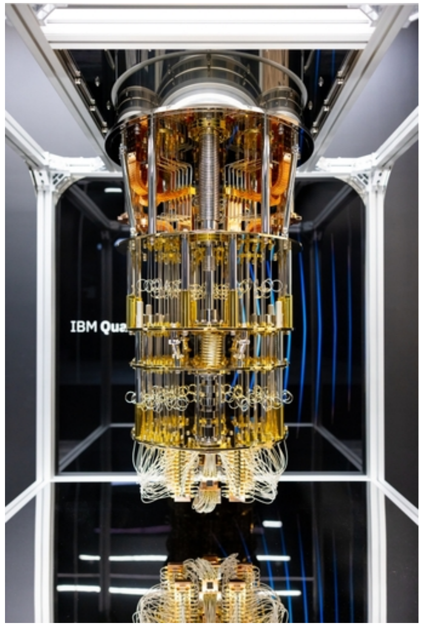
technology-independent

...





“data mechanic”



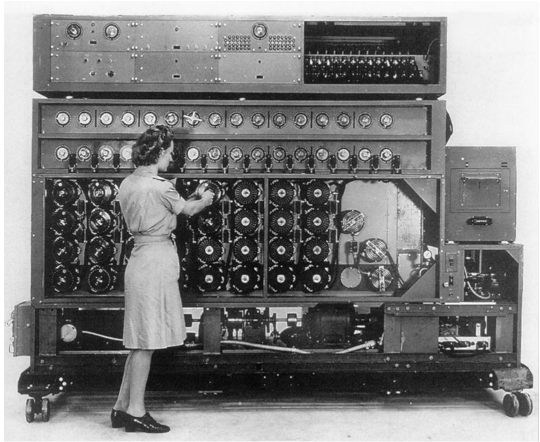
?

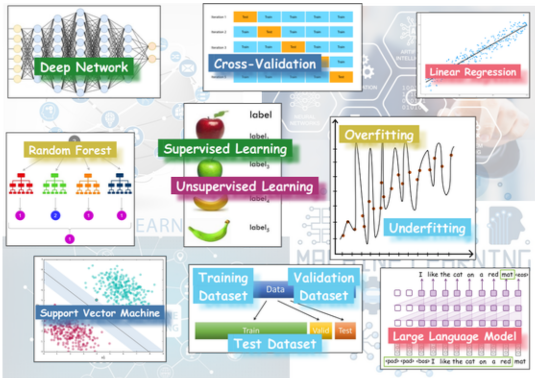
cartwright

car mechanic

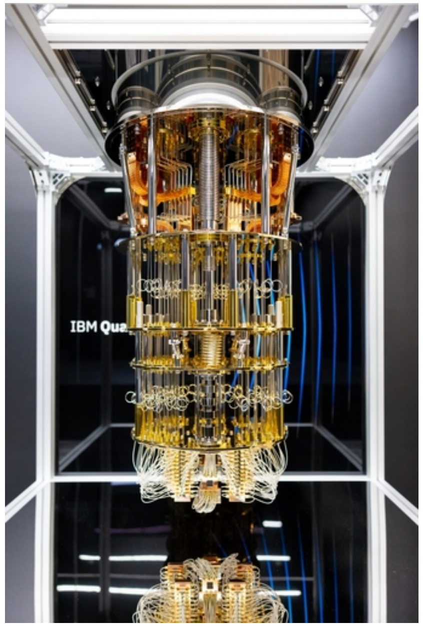
EV mechanic

?





"data mechanic"



?

"data engineer"

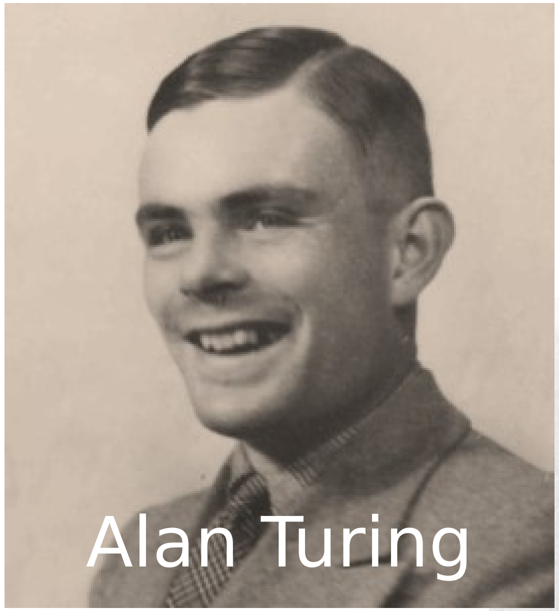
[insert picture of you all here?]

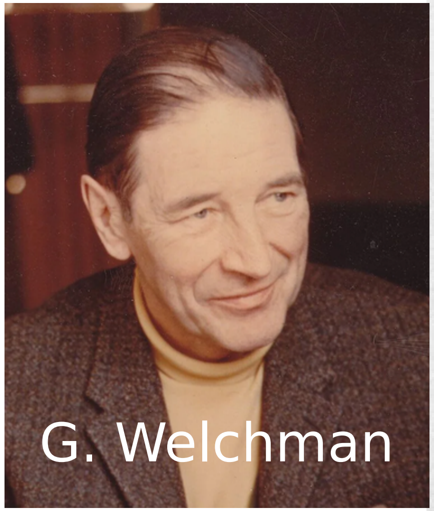
The Bombe a.k.a. *Agnes*

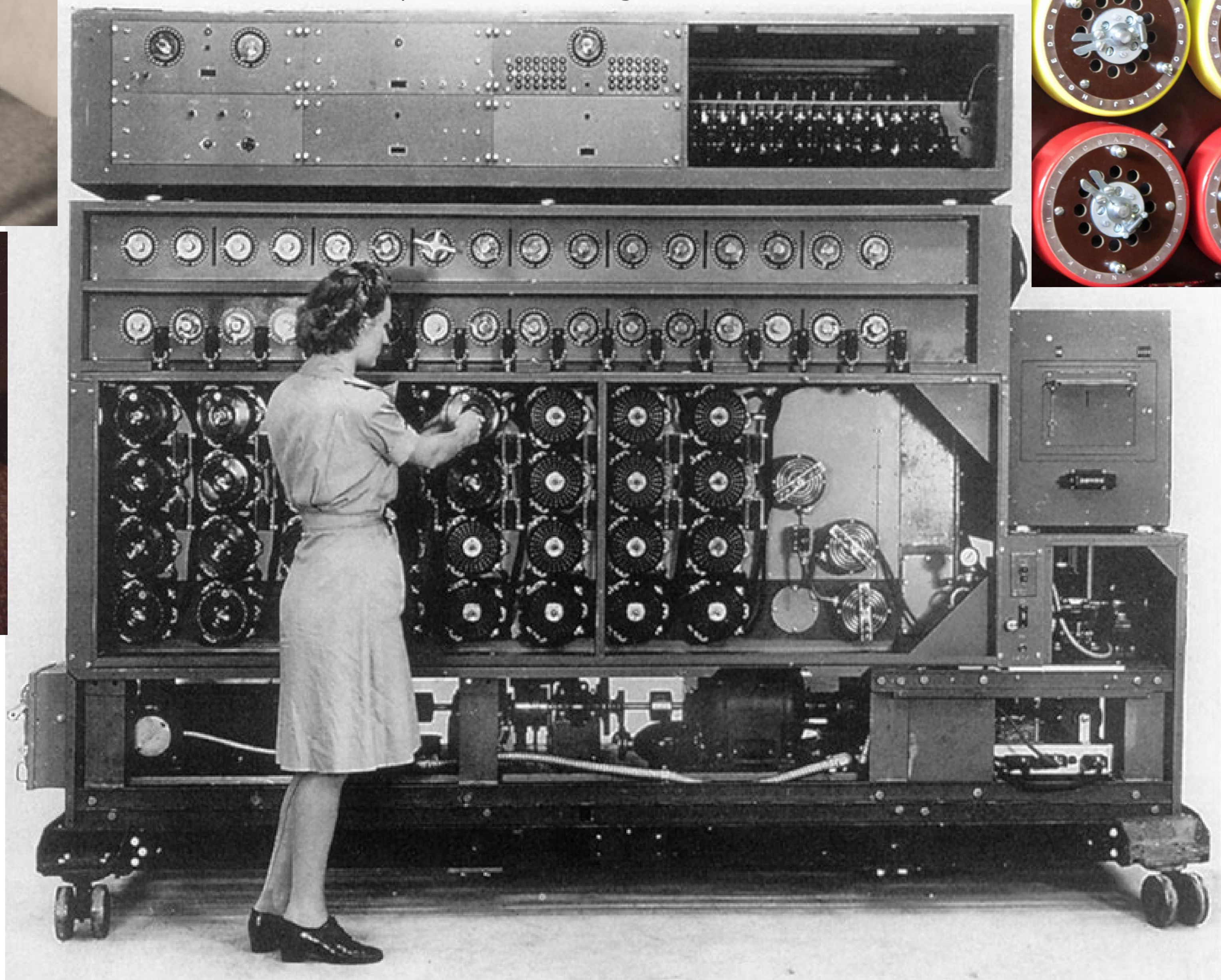
electromechanical computer

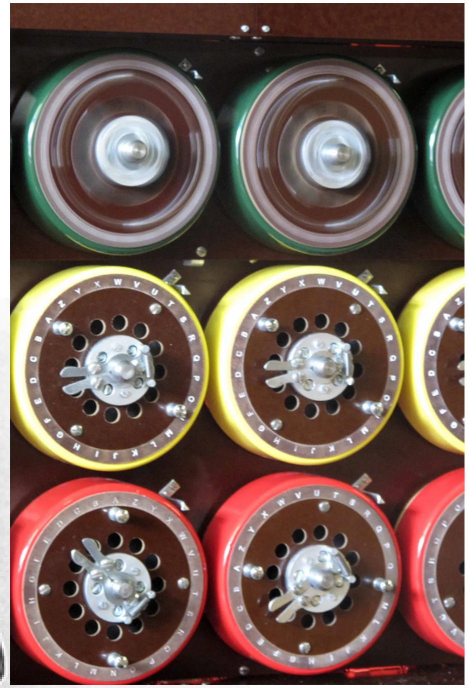
Decoded encrypted Nazi communications during World War II
(efficiency rate: 98.8%)

https://www.britannica.com/topic/Bombe
https://www.tnmoc.org/bombe

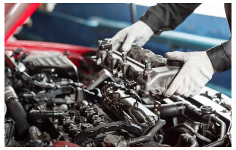


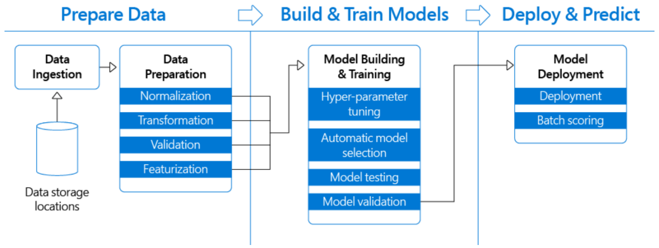




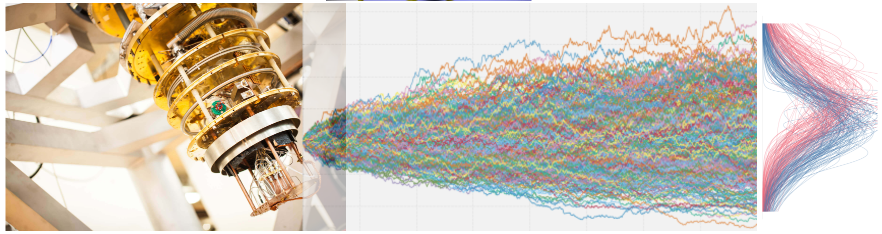


“data mechanic”





“data engineer”



specific

practical

procedural

technology-dependent

...

knowledge

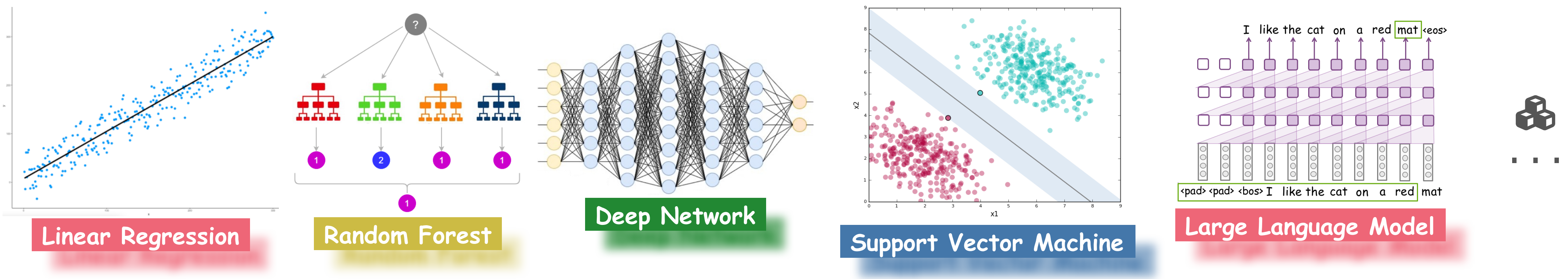
general

principled

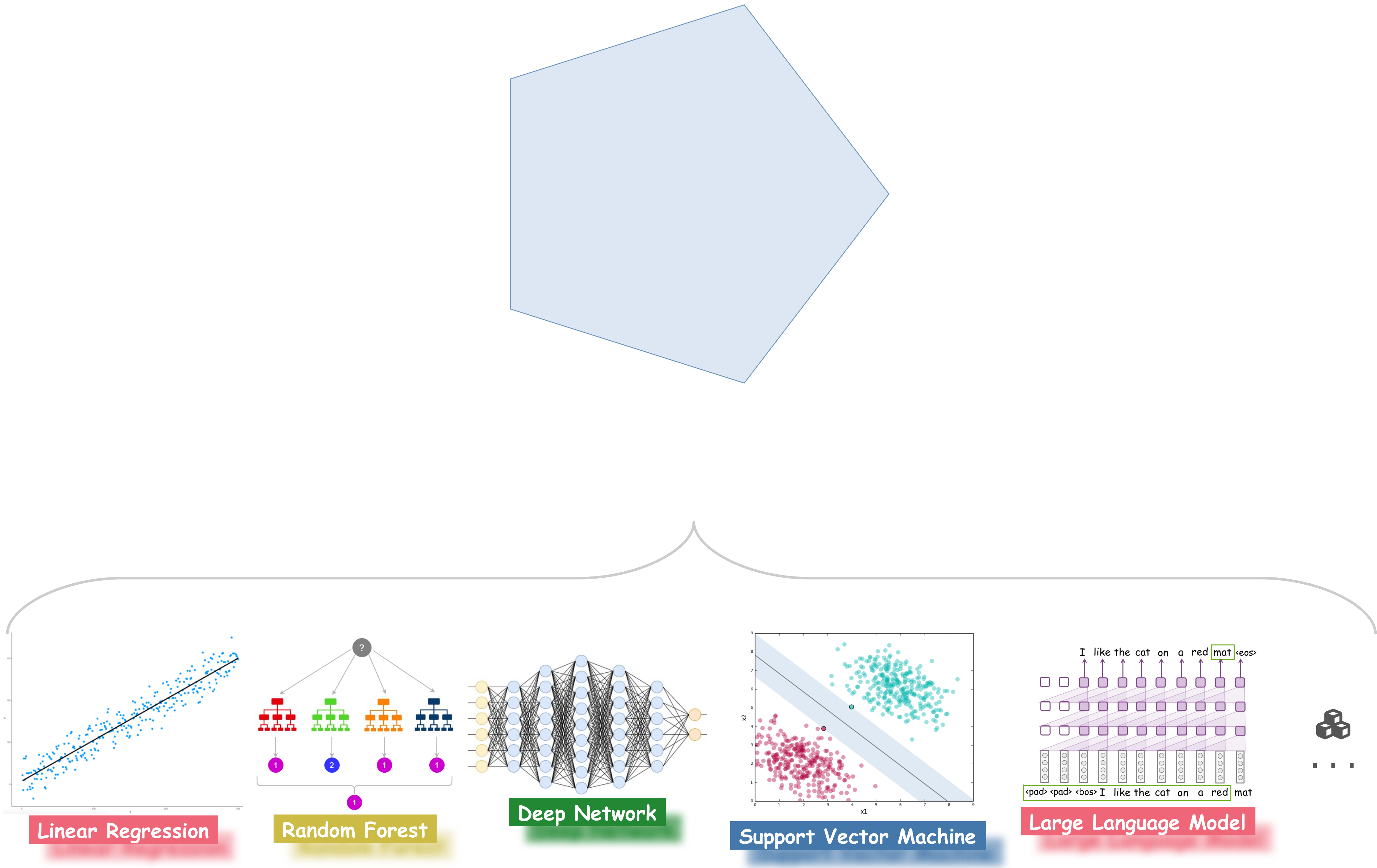
exploratory

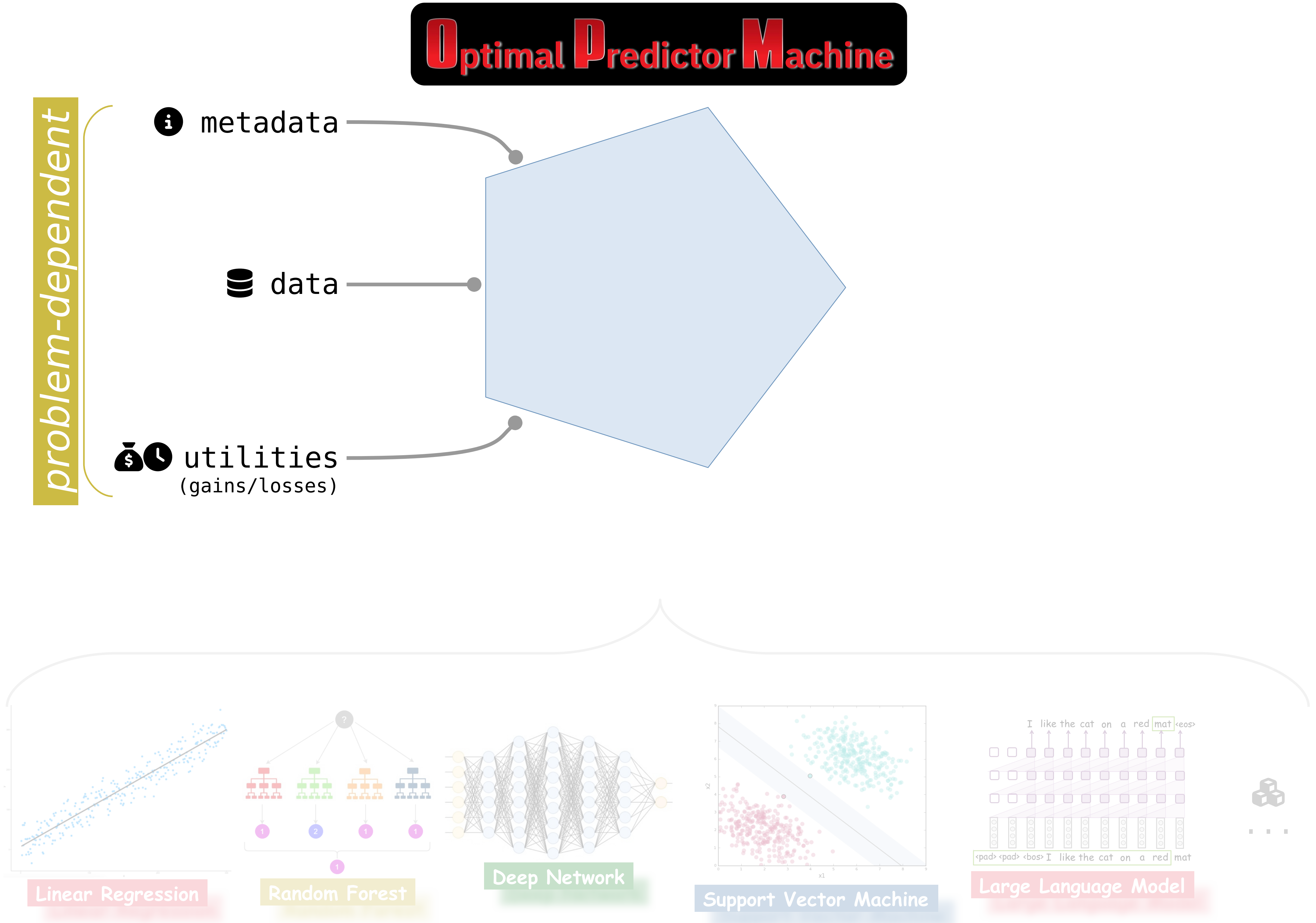
technology-independent

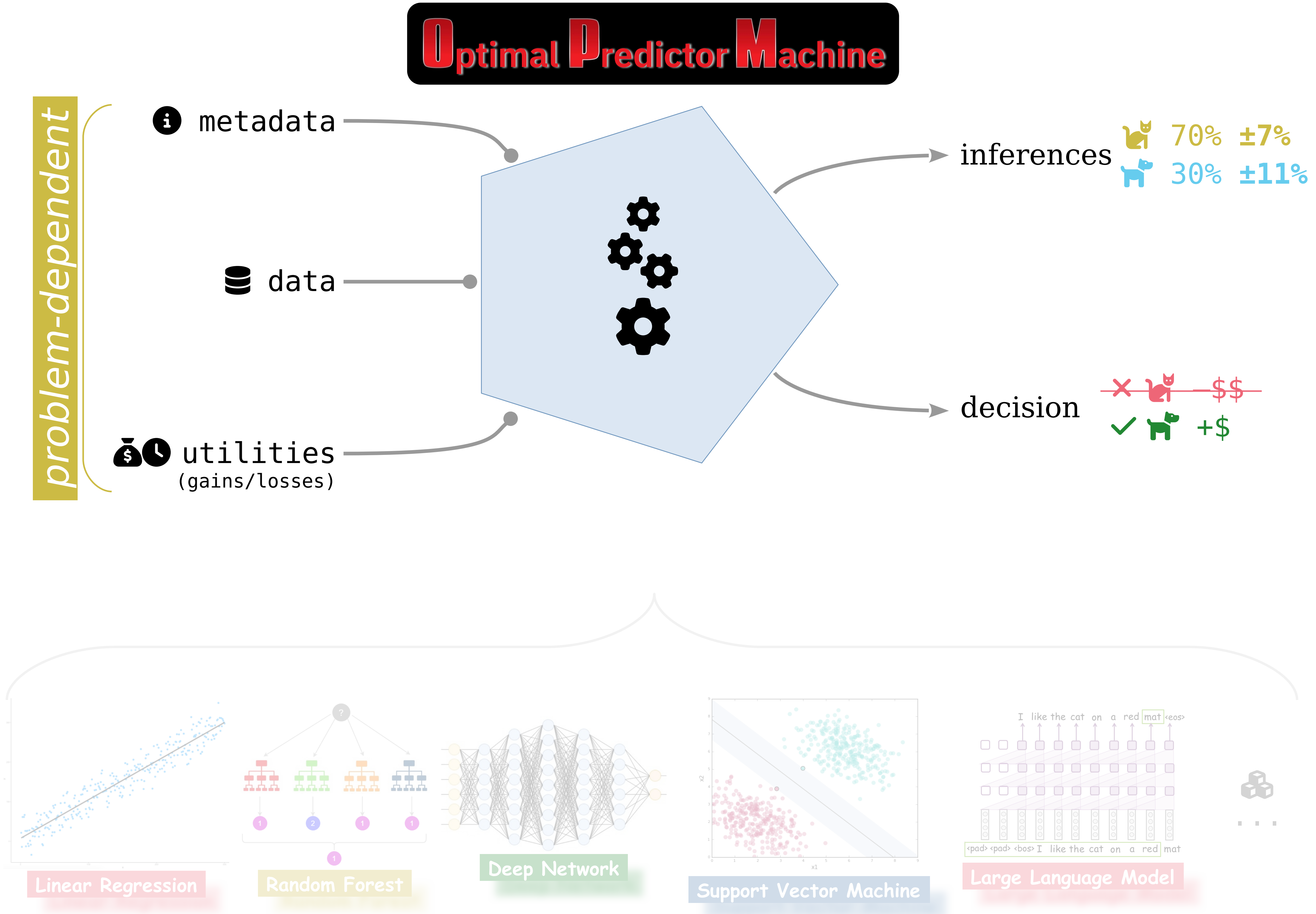
...


Linear Regression
Random Forest
Deep Network
Support Vector Machine
Large Language Model
I like the cat on a red mat <eos>
<pad> <pad> <bos> I like the cat on a red mat
x1
x2
...

Optimal Predictor Machine

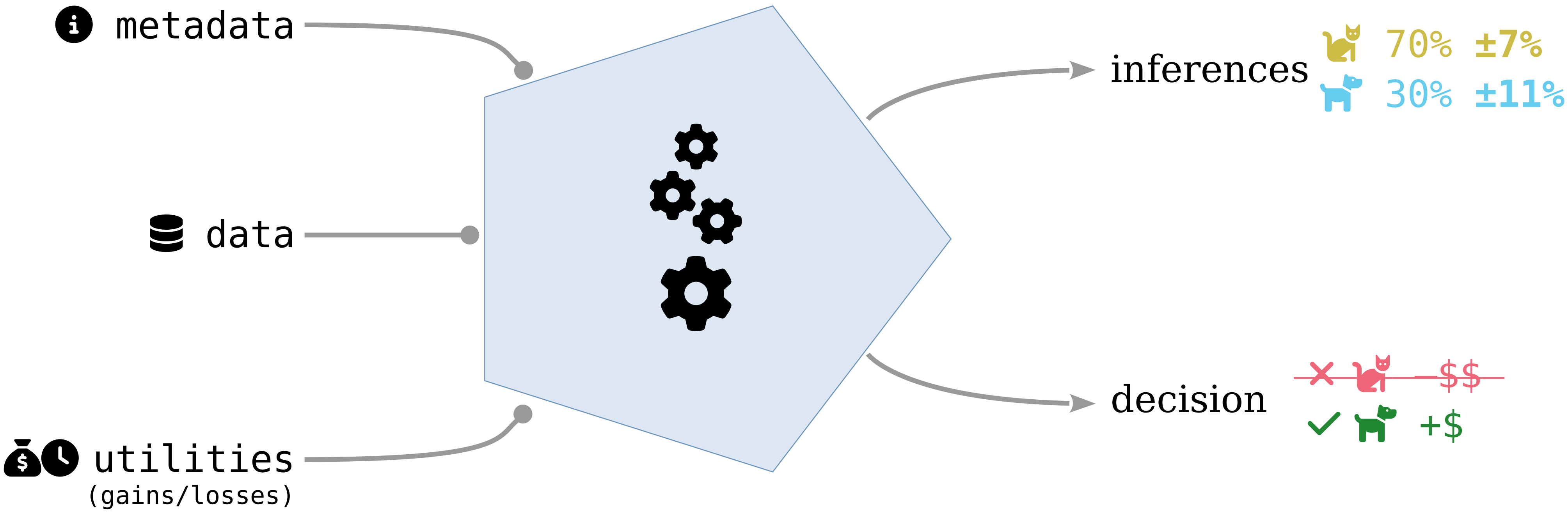



Optimal Predictor Machine
problem-dependent
metadata
data
utilities
(gains/losses)
Linear Regression
Random Forest
Deep Network
Support Vector Machine
Large Language Model
I like the cat on a red mat <eos>
<pad> <pad> <bos> I like the cat on a red mat
...



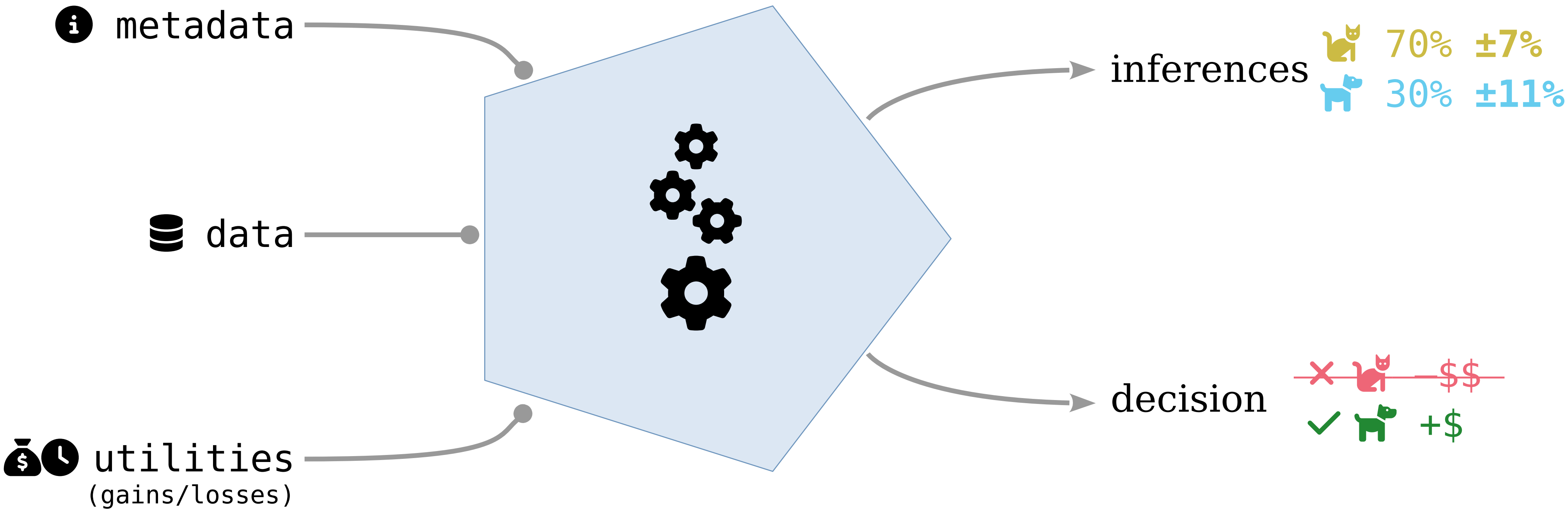
Optimal Predictor Machine
problem-dependent
metadata
data
utilities
(gains/losses)
inferences
70% ±7%
30% ±11%
decision
−$$
+$
Linear Regression
Random Forest
Deep Network
Support Vector Machine
Large Language Model

Optimal Predictor Machine

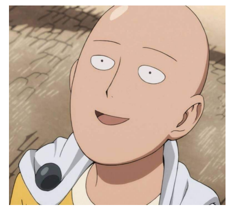


✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!

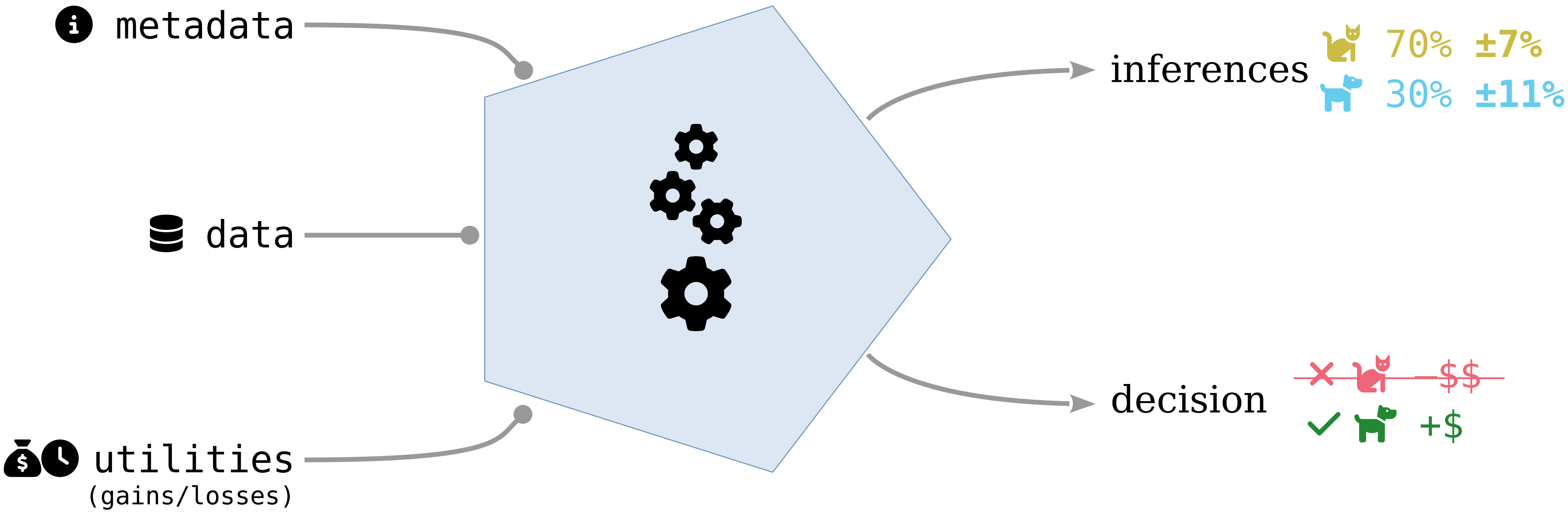
Optimal Predictor Machine



✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!

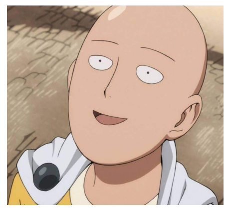

(like One Punch Man)

Optimal Predictor Machine



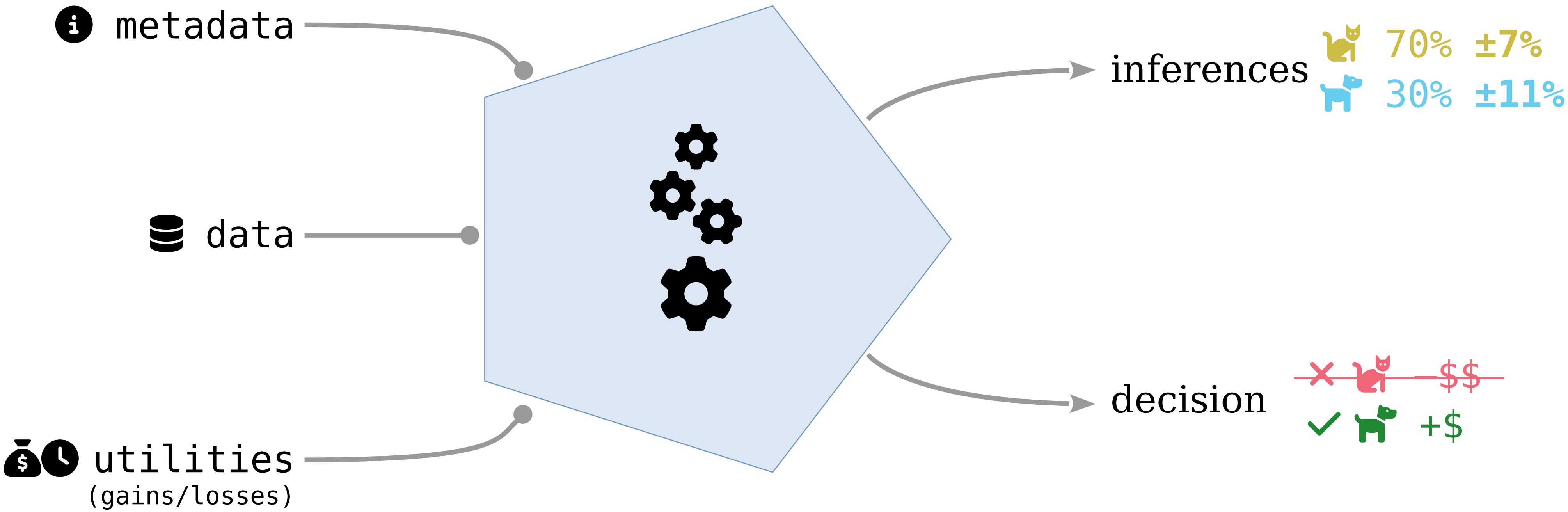
✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!

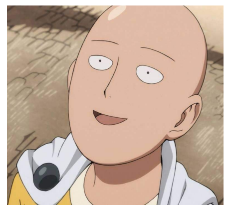
✓ mathematically guaranteed maximally-informative output



(like One Punch Man)

Optimal Predictor Machine



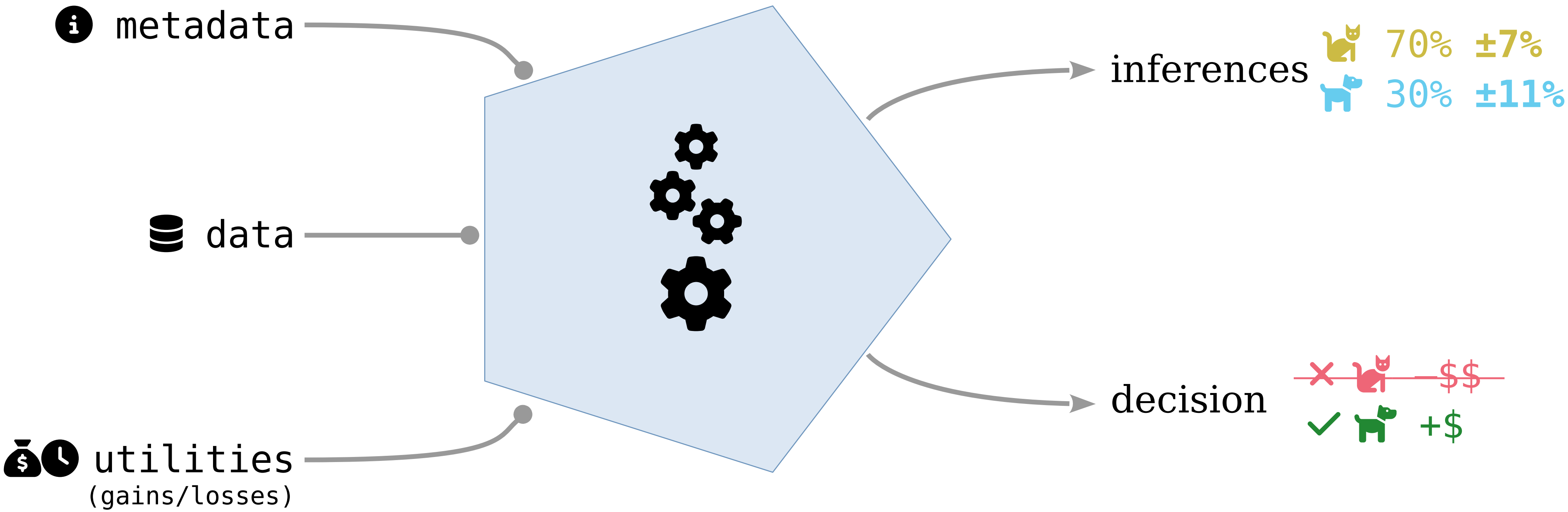

(like One Punch Man)

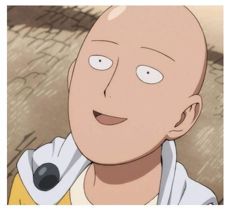
✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!

✓ mathematically guaranteed maximally-informative output

✓ learning is built-in

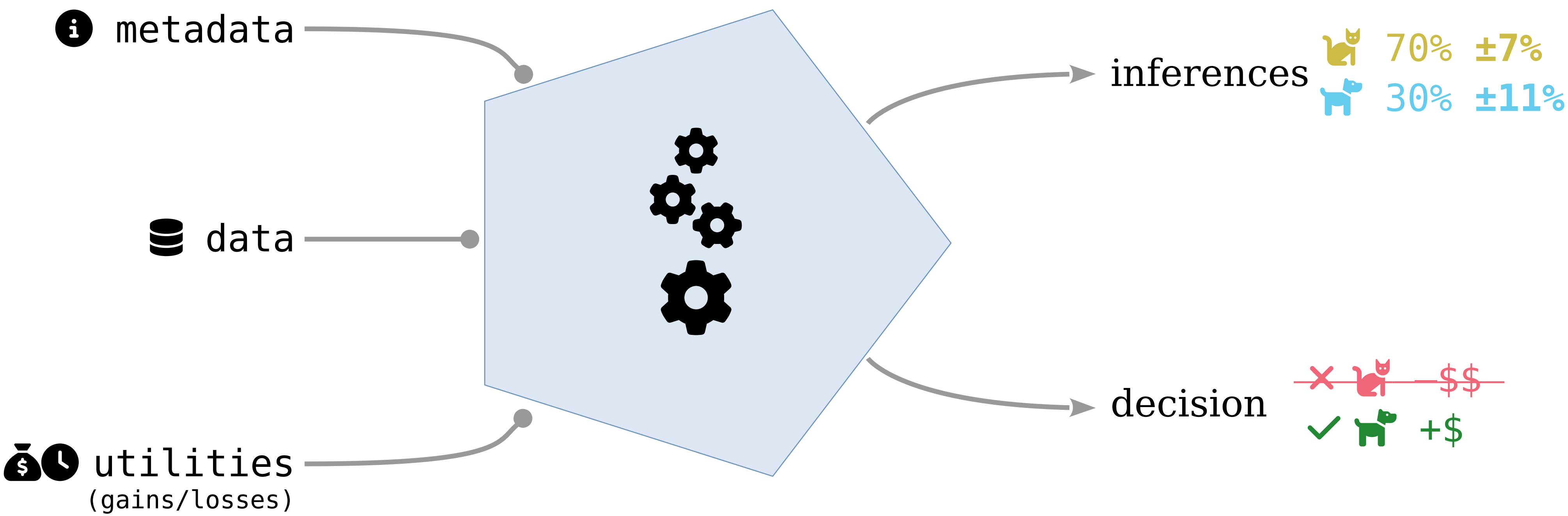
Optimal Predictor Machine




(like One Punch Man)

- ✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!
- ✓ mathematically guaranteed maximally-informative output
- ✓ learning is built-in
- ✓ no need to handle “un/supervised” or “discriminative/generative” differently

Optimal Predictor Machine



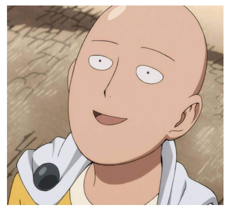
✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!

✓ mathematically guaranteed maximally-informative output

✓ learning is built-in

✓ no need to handle “un/supervised” or “discriminative/generative” differently

✓ no need for data splits, cross-validation, under/overfit... (all implicit in how it works)



(like One Punch Man)

$$P(\neg Y \mid X) = 1 - P(Y \mid X)$$

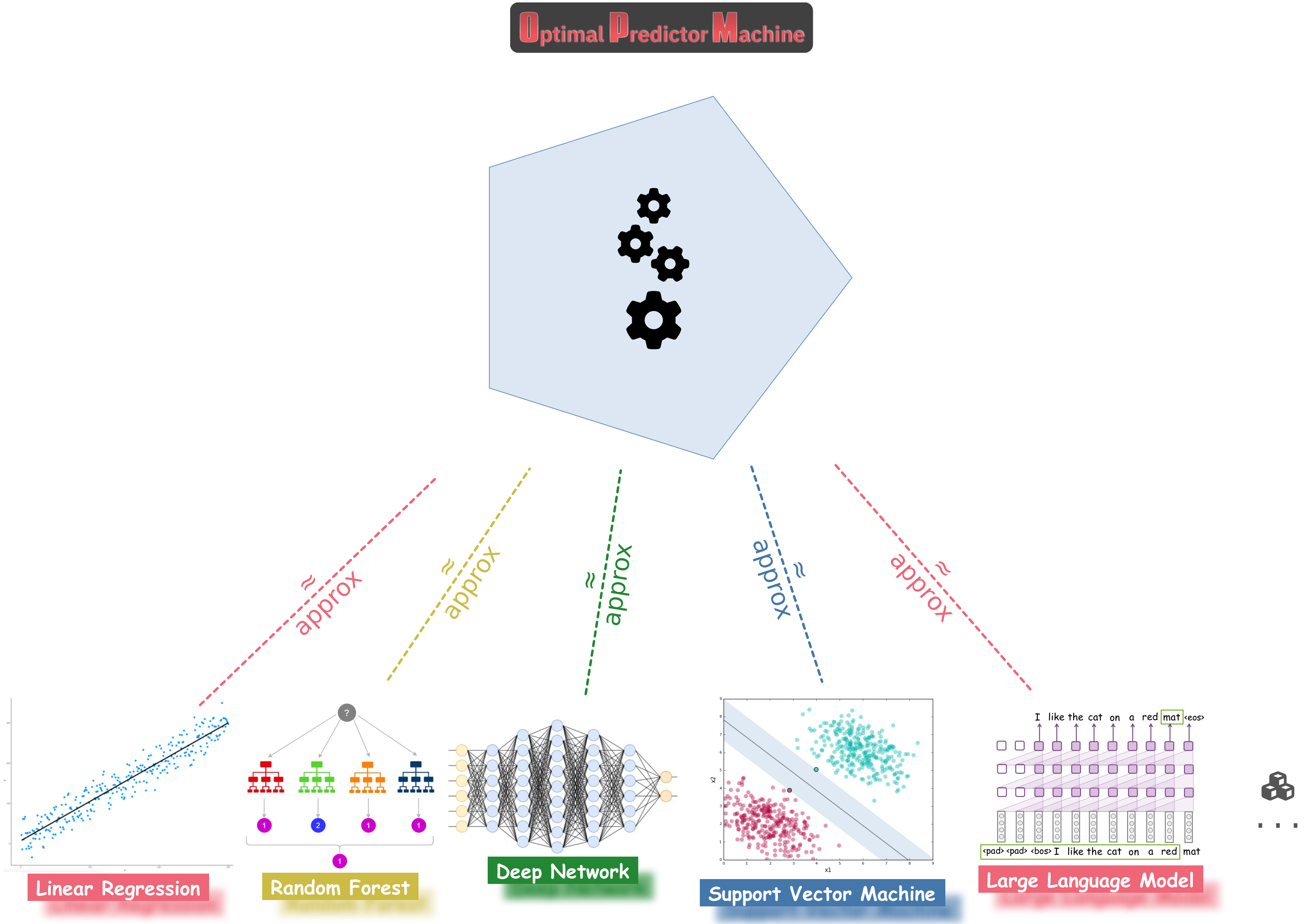
$$P(Z \wedge Y \mid X) = P(Z \mid Y \wedge X) \cdot P(Y \mid X)$$

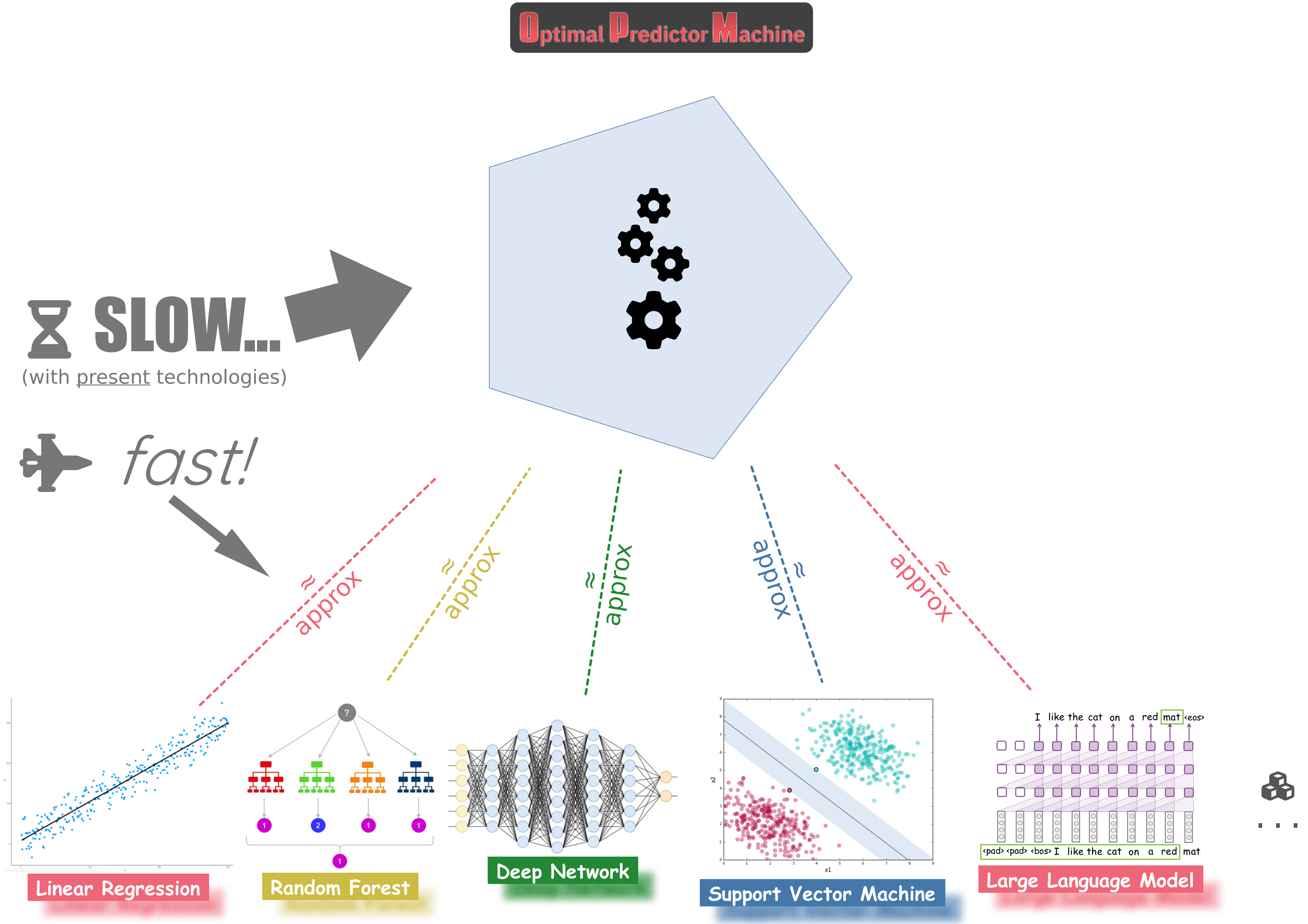
$$P(Z \vee Y \mid X) = P(Z \mid X) + P(Y \mid X) - P(Z \wedge Y \mid X)$$

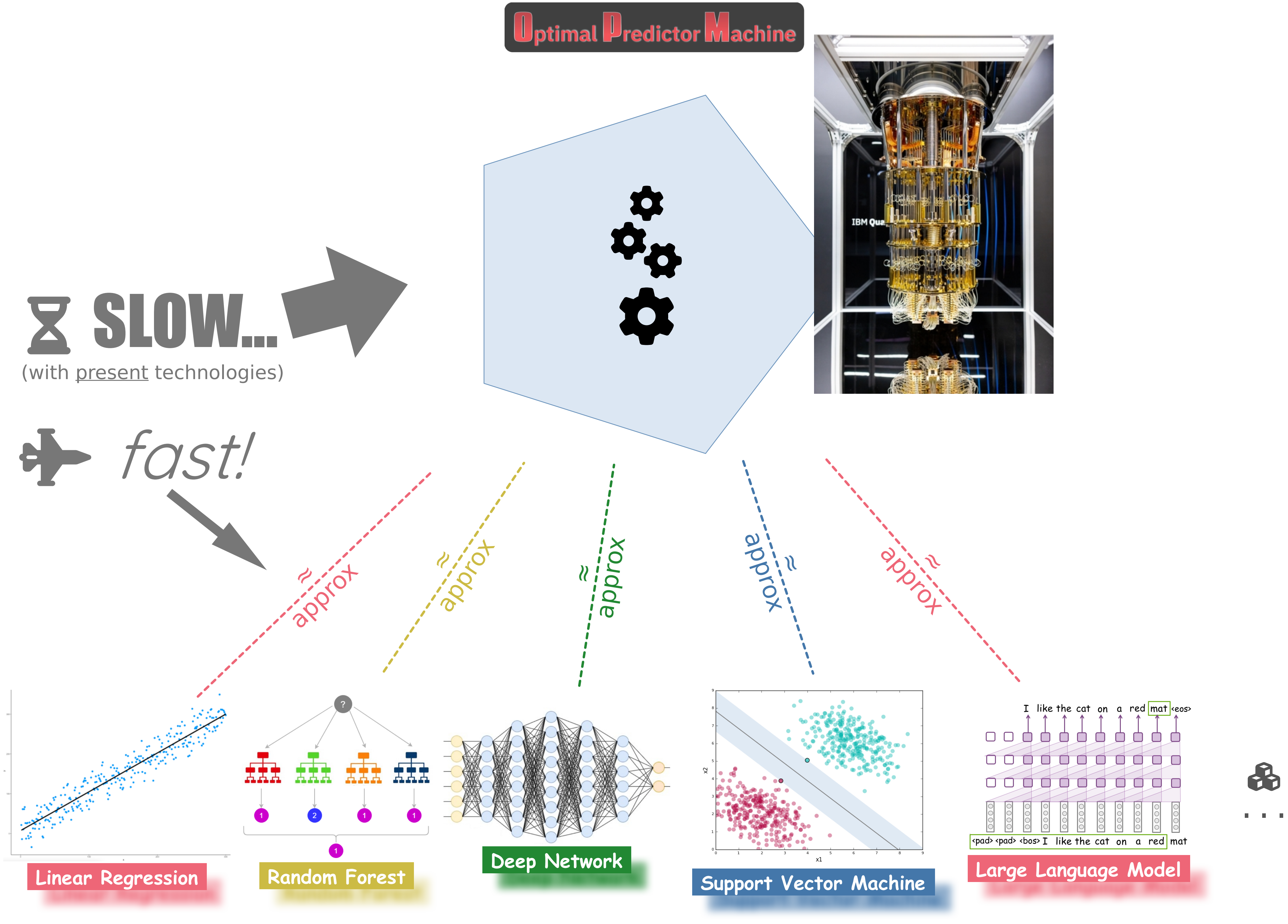
$$\text{choose} \quad \underset{D}{\arg\max}\ U(D \mid Y \wedge X) \cdot P(Y \mid D \wedge X)$$

the machine works by **4 fundamental mechanisms**

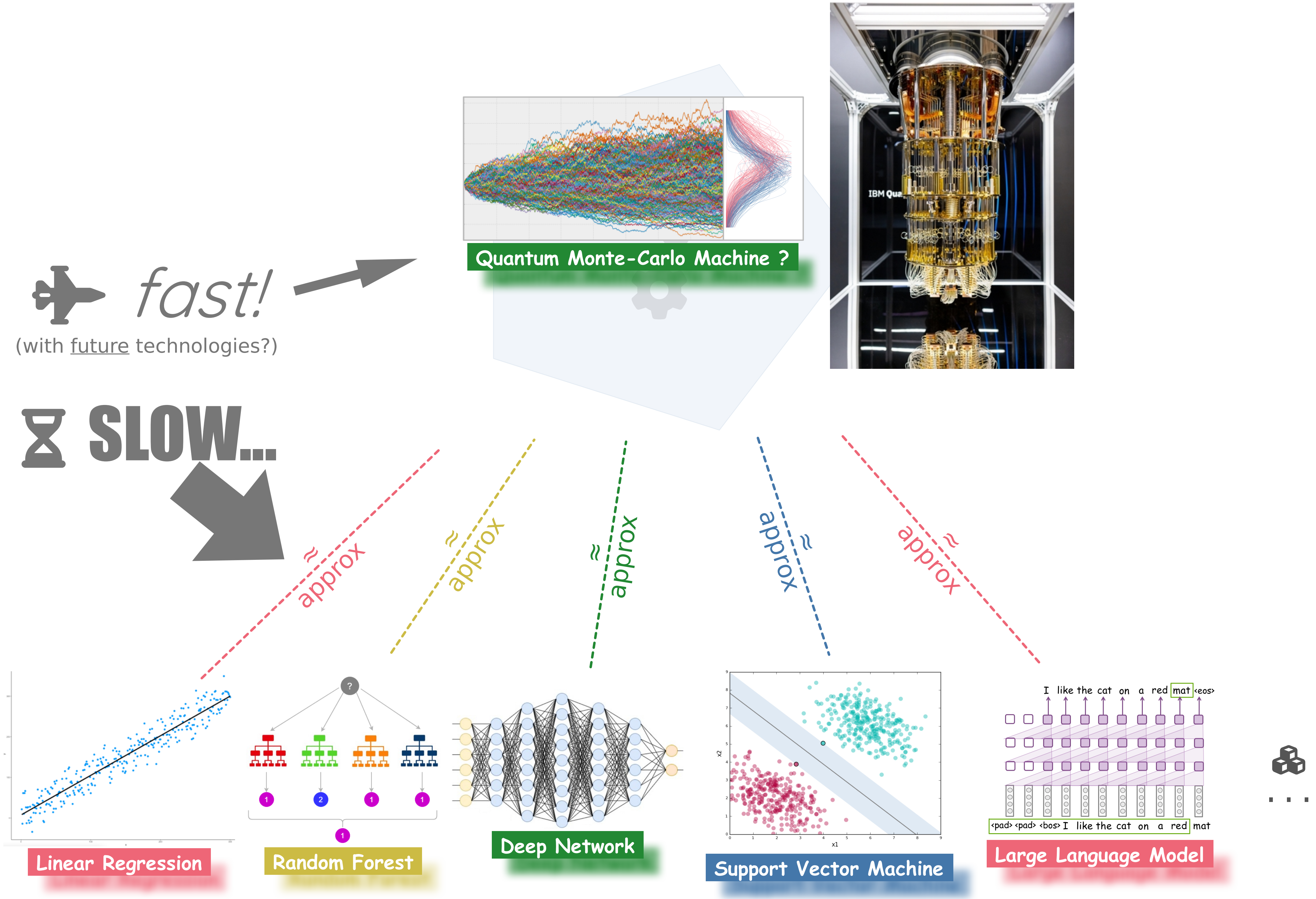
“laws of robotics”
for AI dealing with *uncertainty* & *decision-making*

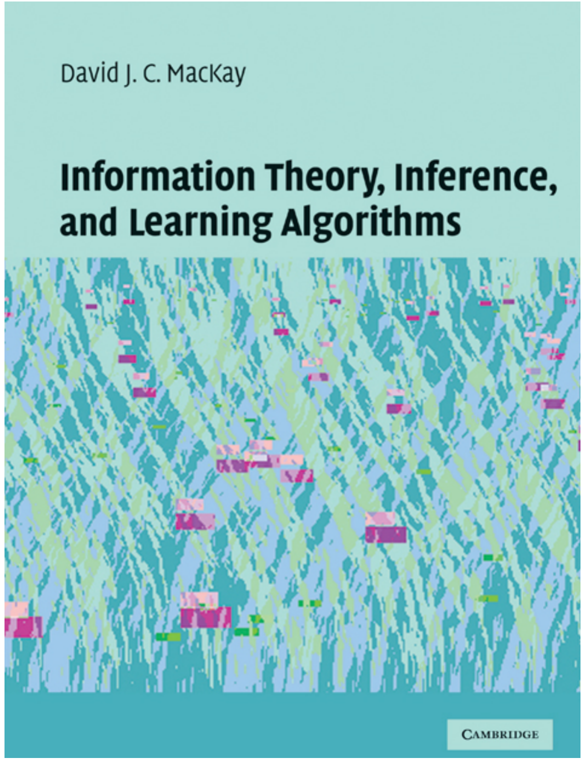

Optimal Predictor Machine
≈ approx
≈ approx
≈ approx
≈ approx
≈ approx
Linear Regression
Random Forest
Deep Network
Support Vector Machine
I like the cat on a red mat <eos>
<pad> <pad> <bos> I like the cat on a red mat
Large Language Model
...

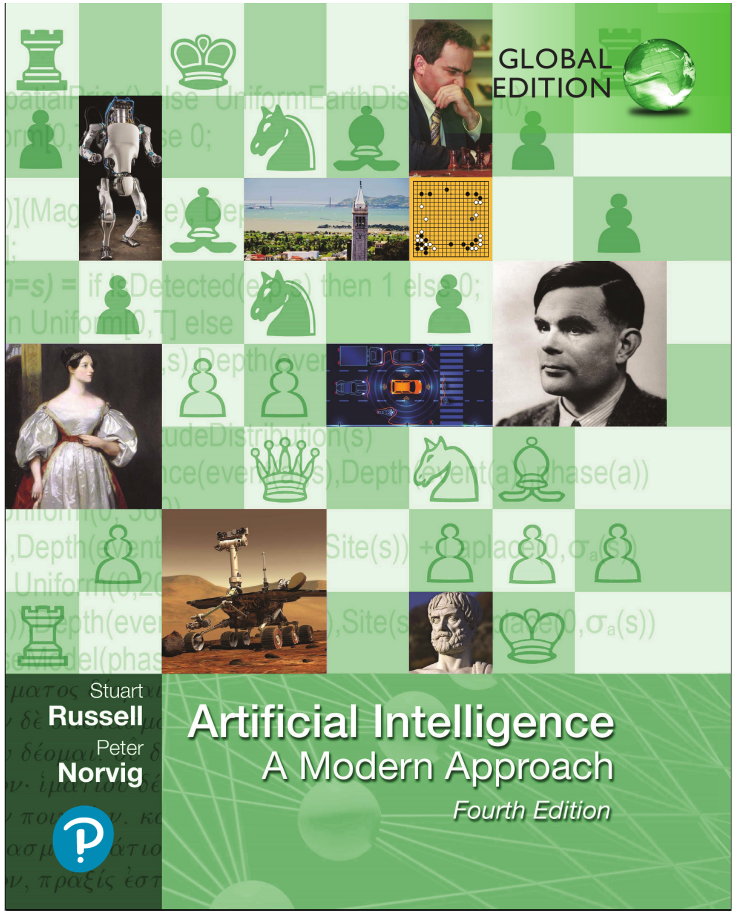

Optimal Predictor Machine
SLOW...
(with present technologies)
fast!
≈ approx
≈ approx
≈ approx
≈ approx
≈ approx
Linear Regression
Random Forest
Deep Network
Support Vector Machine
I like the cat on a red mat <eos>
<pad> <pad> <bos> I like the cat on a red mat
Large Language Model
...

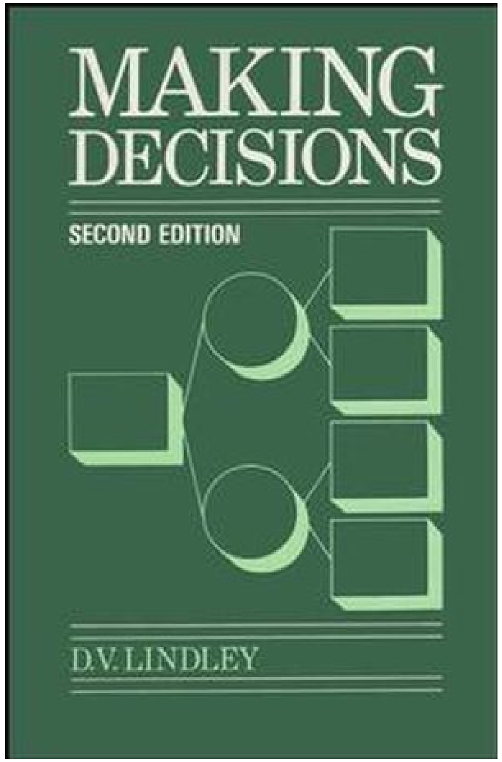


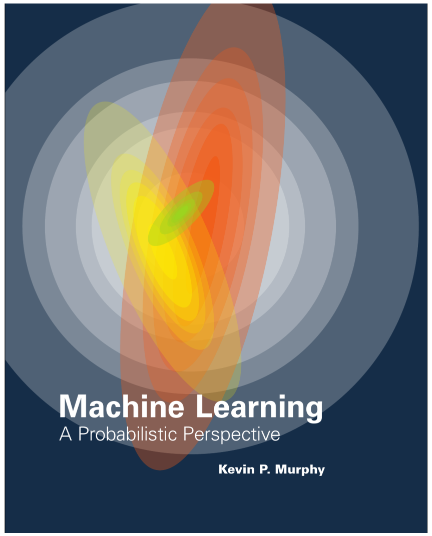
Optimal Predictor Machine
IBM Qua
SLOW...
(with present technologies)
fast!
≈ approx
≈ approx
≈ approx
≈ approx
≈ approx
Linear Regression
Random Forest
Deep Network
Support Vector Machine
Large Language Model
I like the cat on a red mat <eos>
<pad> <pad> <bos> I like the cat on a red mat
x1
x2
...


Quantum Monte-Carlo Machine ?
fast!
(with future technologies?)
SLOW...
≈ approx
≈ approx
≈ approx
approx ≈
approx ≈
Linear Regression
Random Forest
Deep Network
Support Vector Machine
I like the cat on a red mat <eos>
<pad> <pad> <bos> I like the cat on a red mat
Large Language Model
...


David J. C. MacKay
Information Theory, Inference, and Learning Algorithms
CAMBRIDGE


GLOBAL EDITION
Stuart Russell
Peter Norvig
Artificial Intelligence
A Modern Approach
Fourth Edition


MAKING DECISIONS
SECOND EDITION
D.V. LINDLEY


Machine Learning
A Probabilistic Perspective
Kevin P. Murphy

– Topic plan (subject to changes) –

Goals & strategy

Inference I (and some AI)

- *1st connection with machine learning*

Data I

Inference II

- *2nd connection with machine learning*

Data II

Utility theory (gains & losses)

Decision-making

Machine learning

… and how to improve it!

Canvas

https://hvl.instructure.com/courses/25074

Course notes

https://pglpm.github.io/ADA511/

Software:

R
https://www.r-project.org/

Python
https://www.python.org/