Prompting

- Follow-up from class 3
- Agenda
  - What is a prompt?
  - GPU setup
  - Prompting exercises

Language models

- “What word comes next, given this input?”

```
What        is        your        __?
```

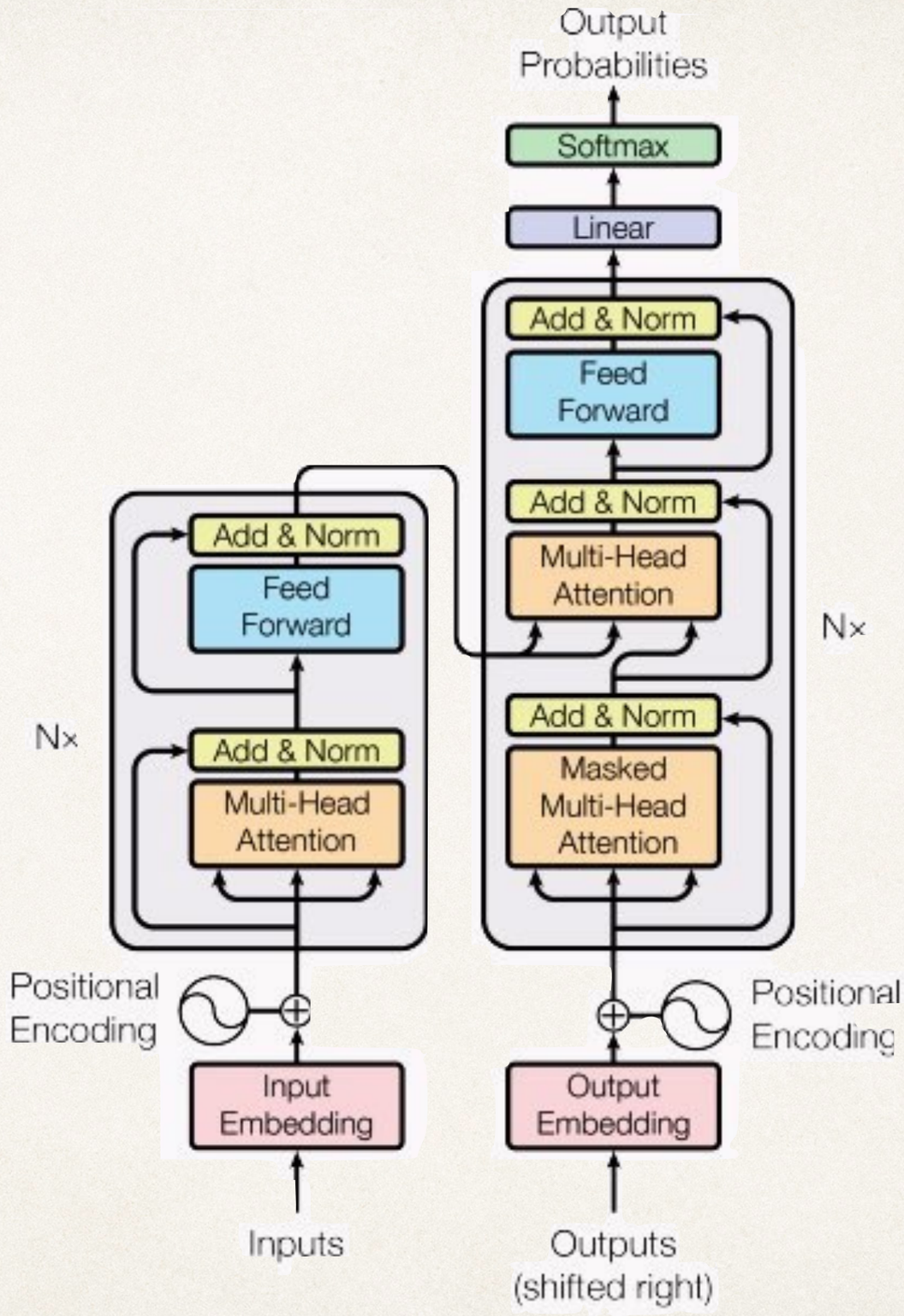
Language models

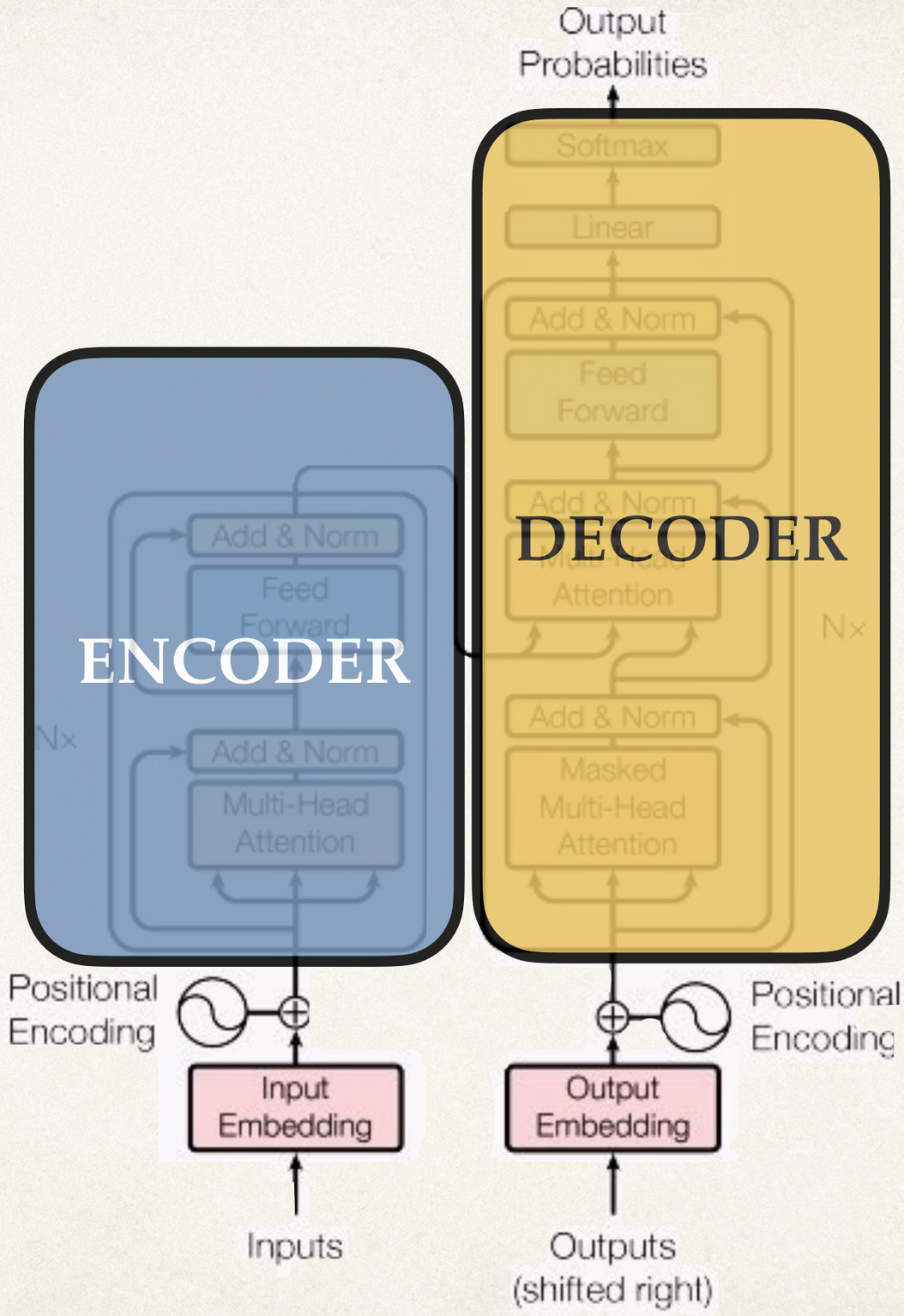
- “What word comes next, given this input?”

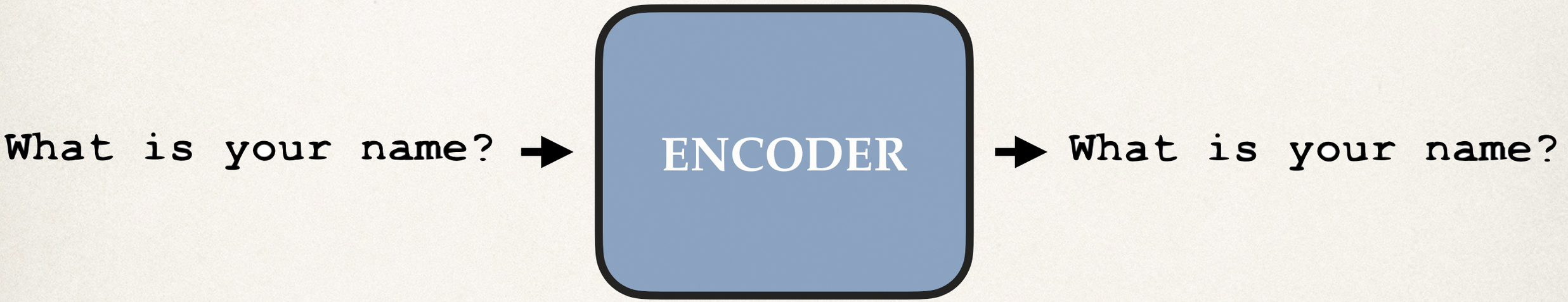
```
What        is        your        __?
```

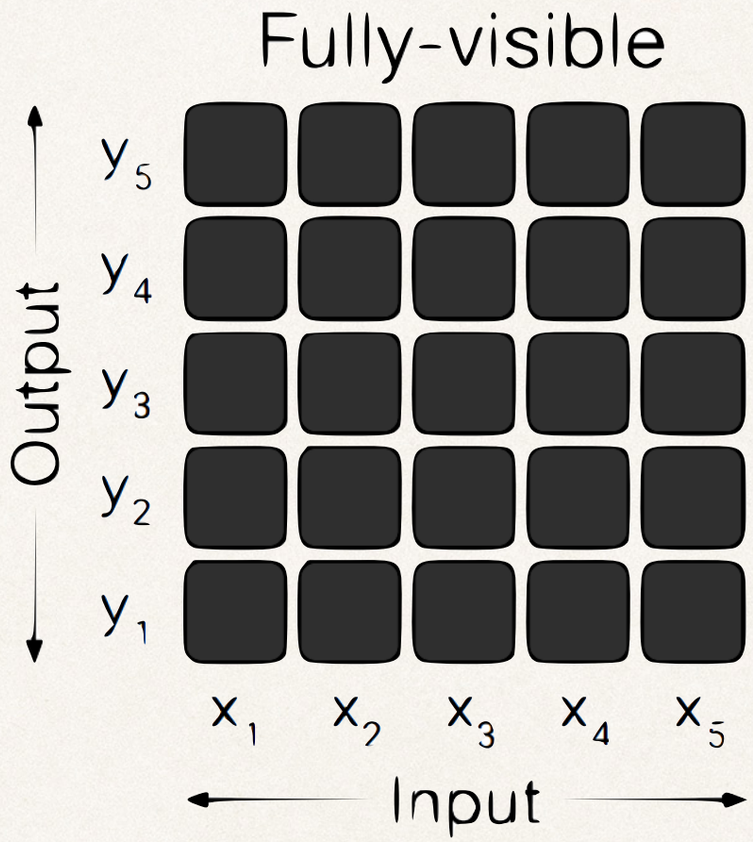
  - Input = **prompt**!

- A type of language model…


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
N×
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
N×
Positional Encoding
Input Embedding
Inputs
Positional Encoding
Output Embedding
Outputs (shifted right)

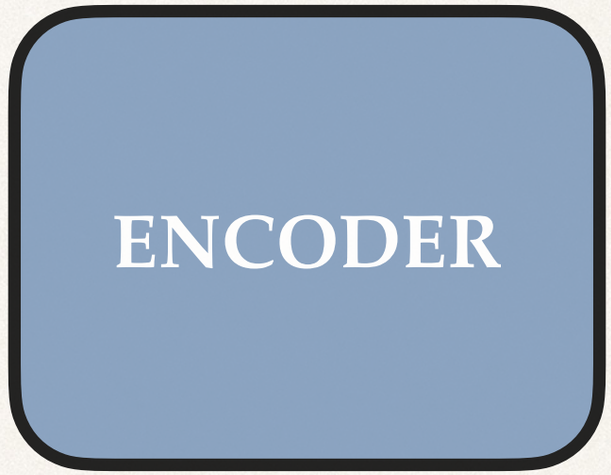

Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Masked Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
DECODER
N×
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
ENCODER
N×
Positional Encoding
Input Embedding
Inputs
Positional Encoding
Output Embedding
Outputs (shifted right)


What is your name?
ENCODER
What is your name?


Fully-visible
y_5
y_4
y_3
y_2
y_1
Output
x_1 x_2 x_3 x_4 x_5
Input

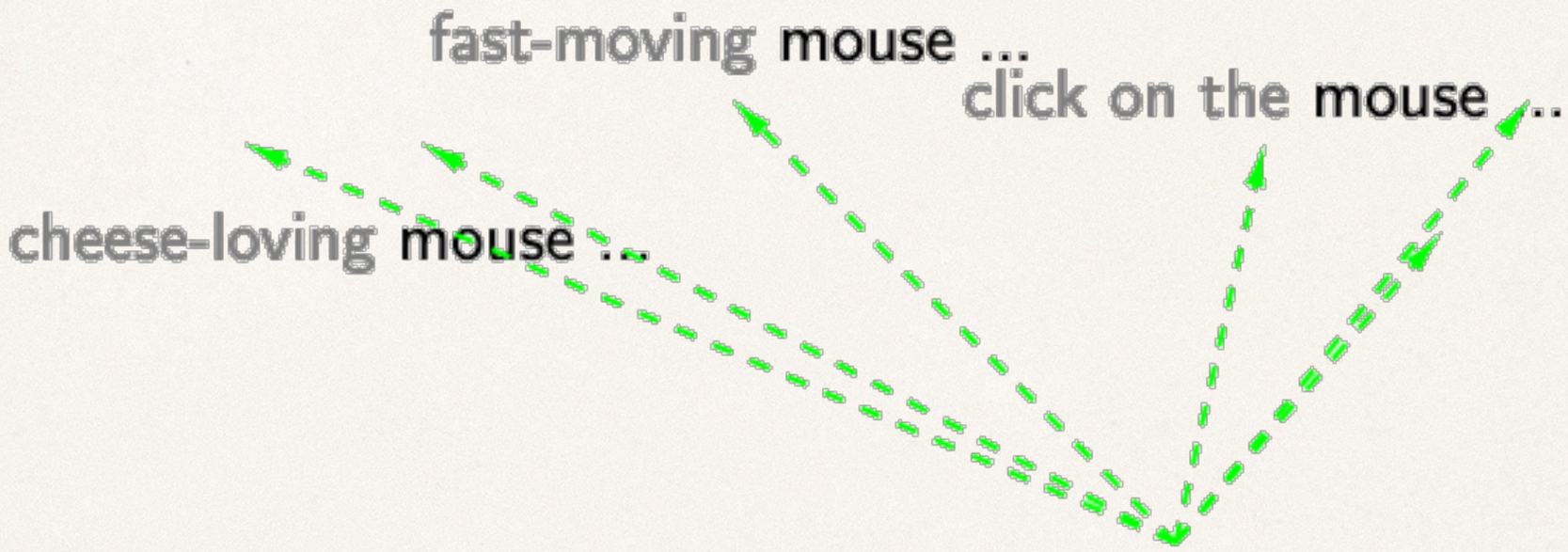
[0.56343, -0.13232 …]

[0.02032, -0.32121 …]



[0.58454, -0.13285 …]

[0.02074, -0.32123 …]


fast-moving mouse ...
click on the mouse ...
cheese-loving mouse ...

Next week!

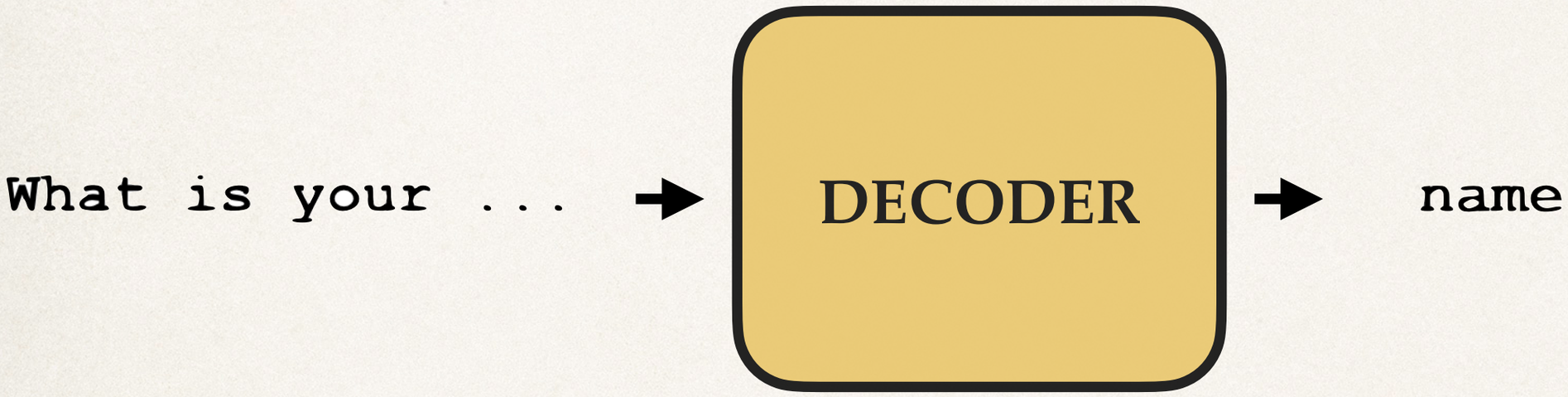
[0.56343, -0.13232 …]

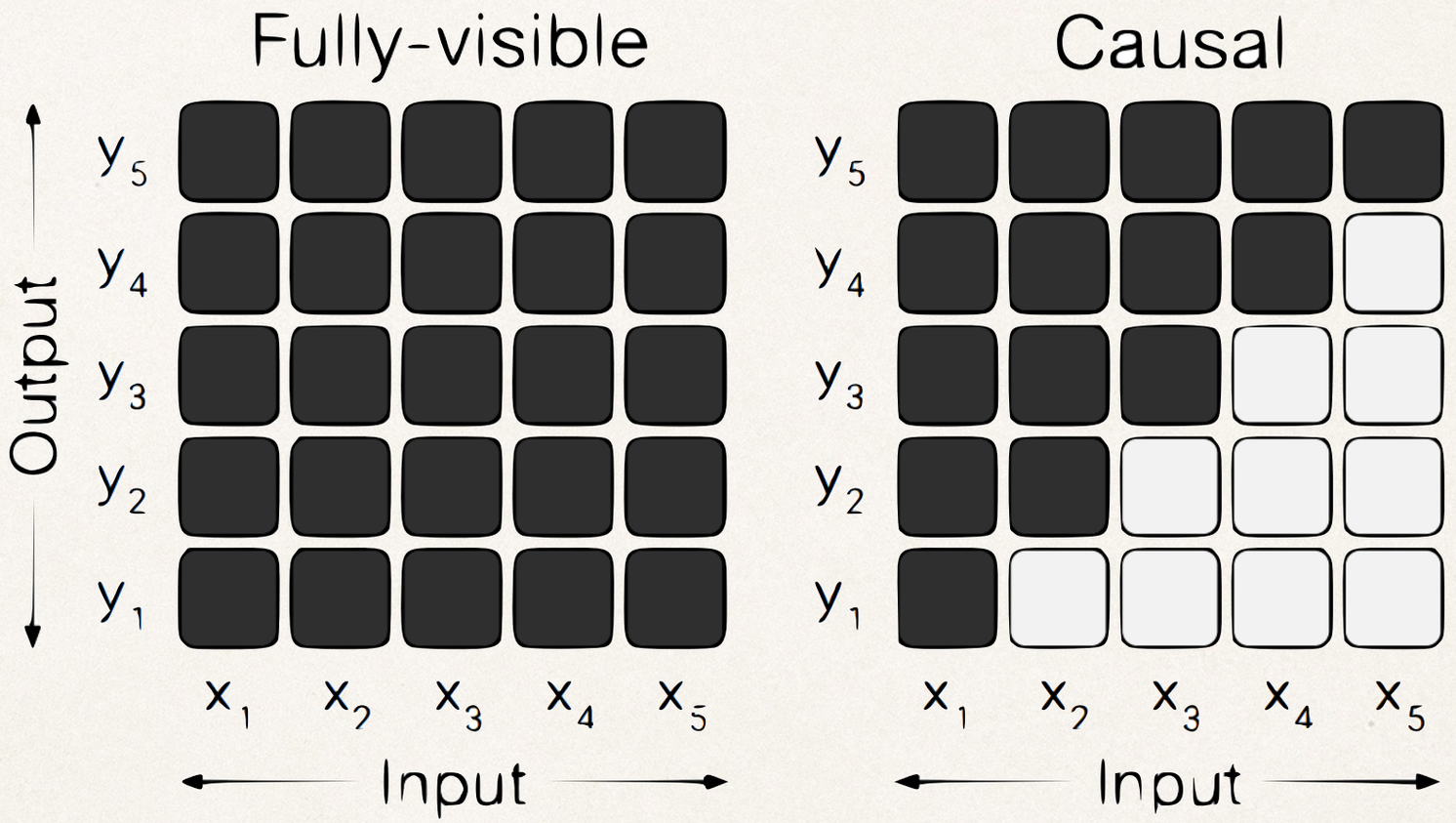
[0.02032, -0.32121 …]

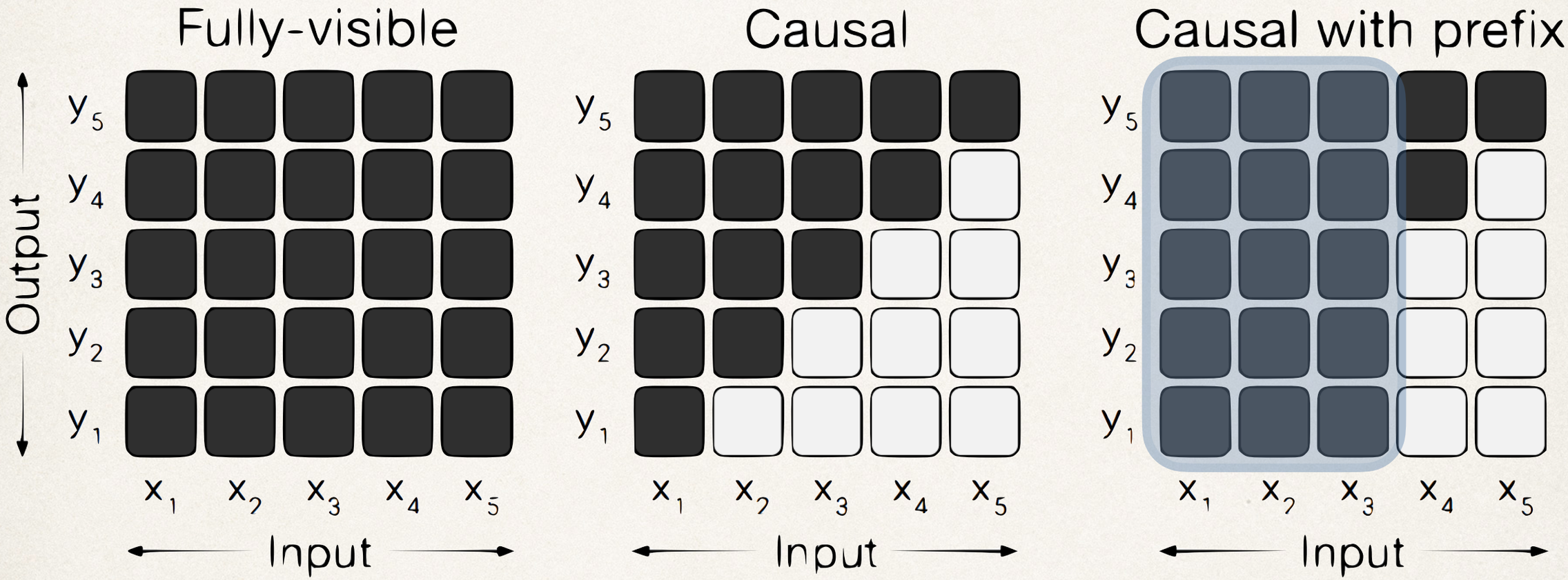
ENCODER

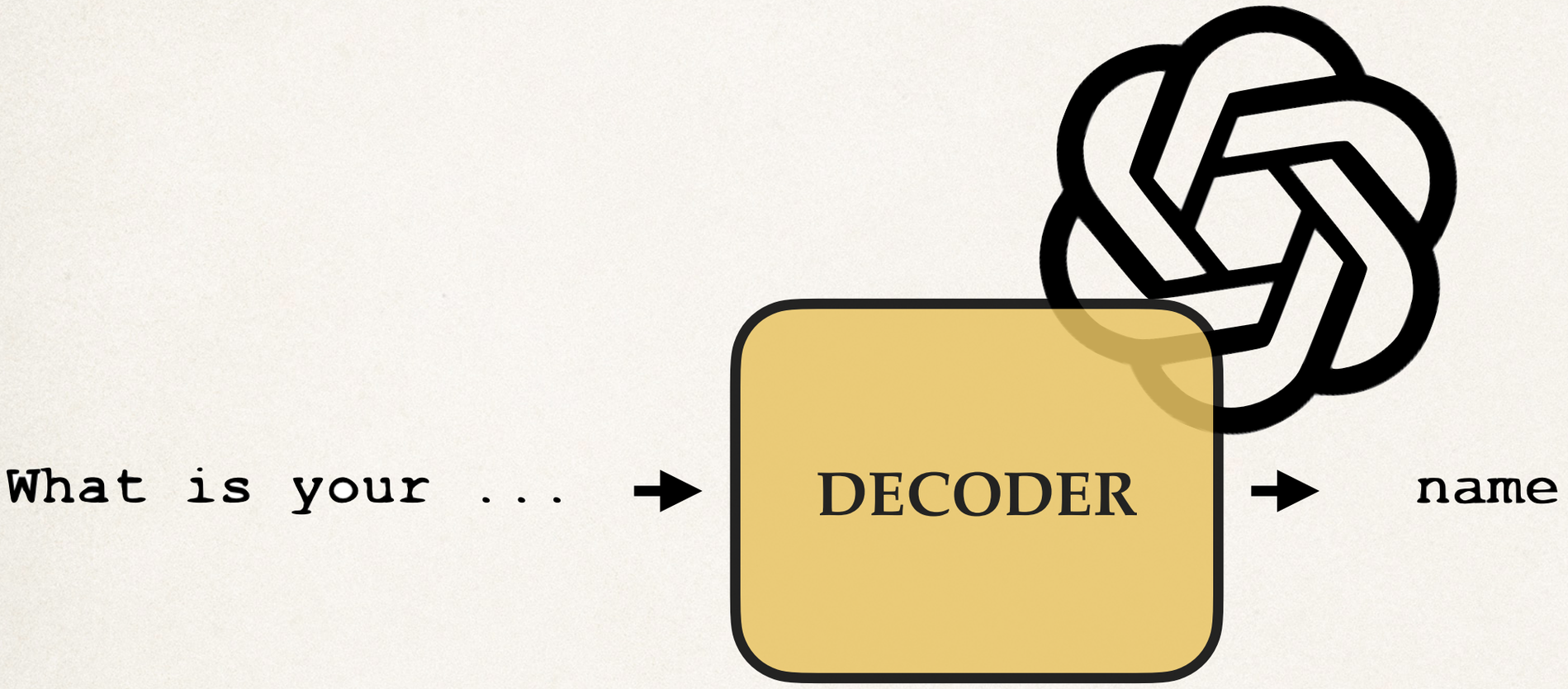
[0.58454, -0.13285 …]

[0.02074, -0.32123 …]

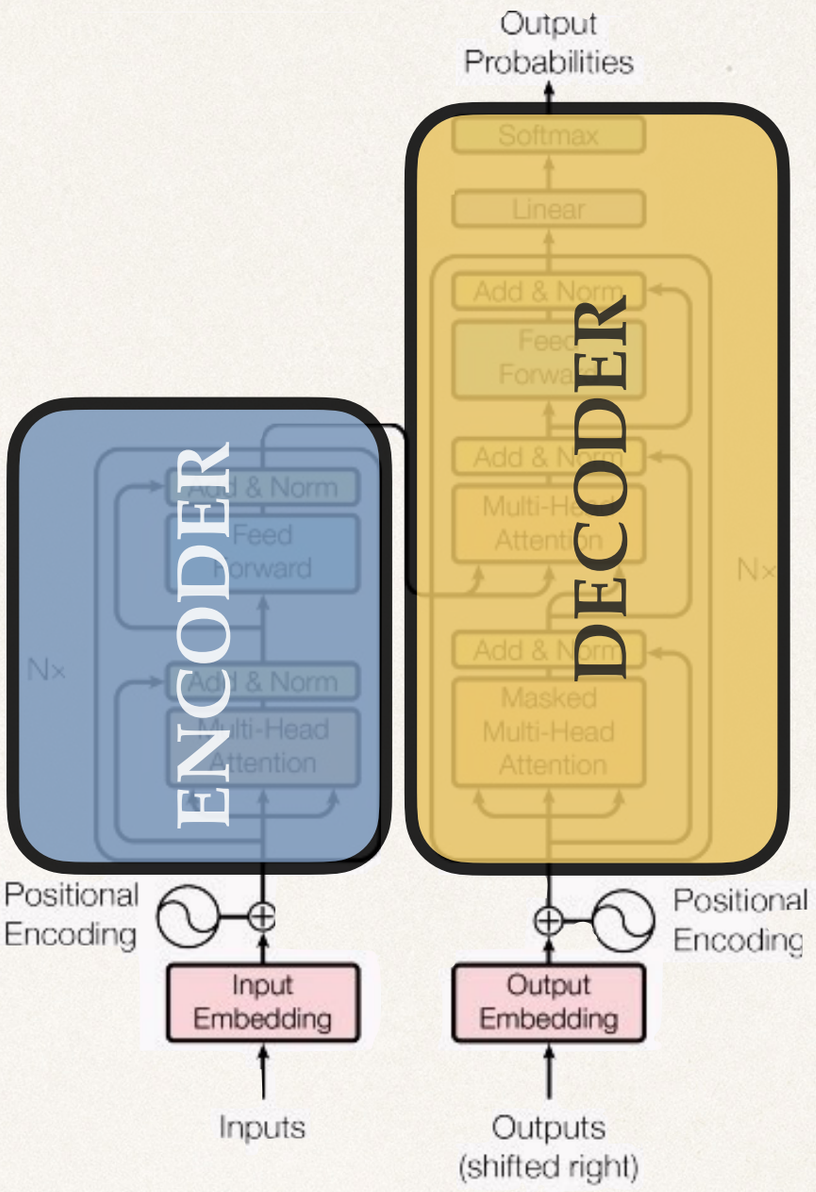

What is your ...
DECODER
name


Fully-visible
Causal
Output
y_5
y_4
y_3
y_2
y_1
x_1 x_2 x_3 x_4 x_5
Input
Input


Fully-visible
Causal
Causal with prefix
Output
y_5
y_4
y_3
y_2
y_1
x_1 x_2 x_3 x_4 x_5
Input

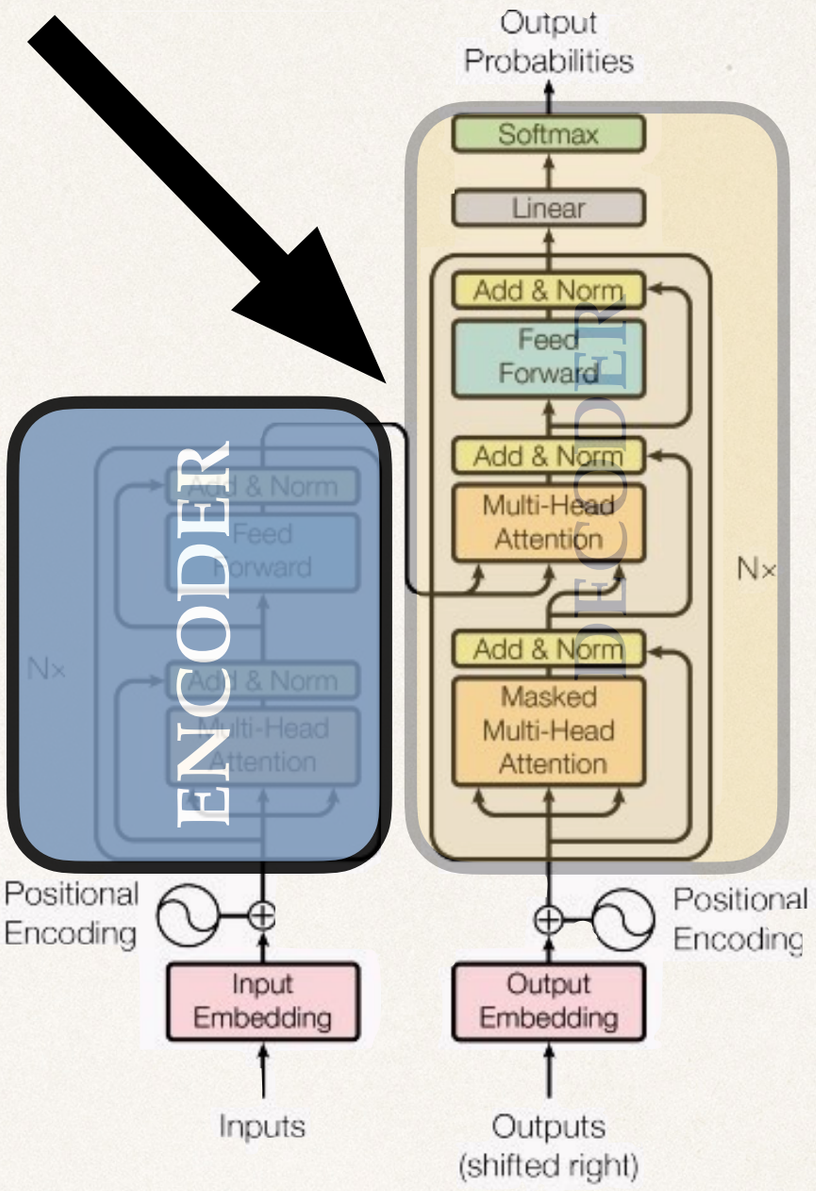

What is your ...
DECODER
name

name

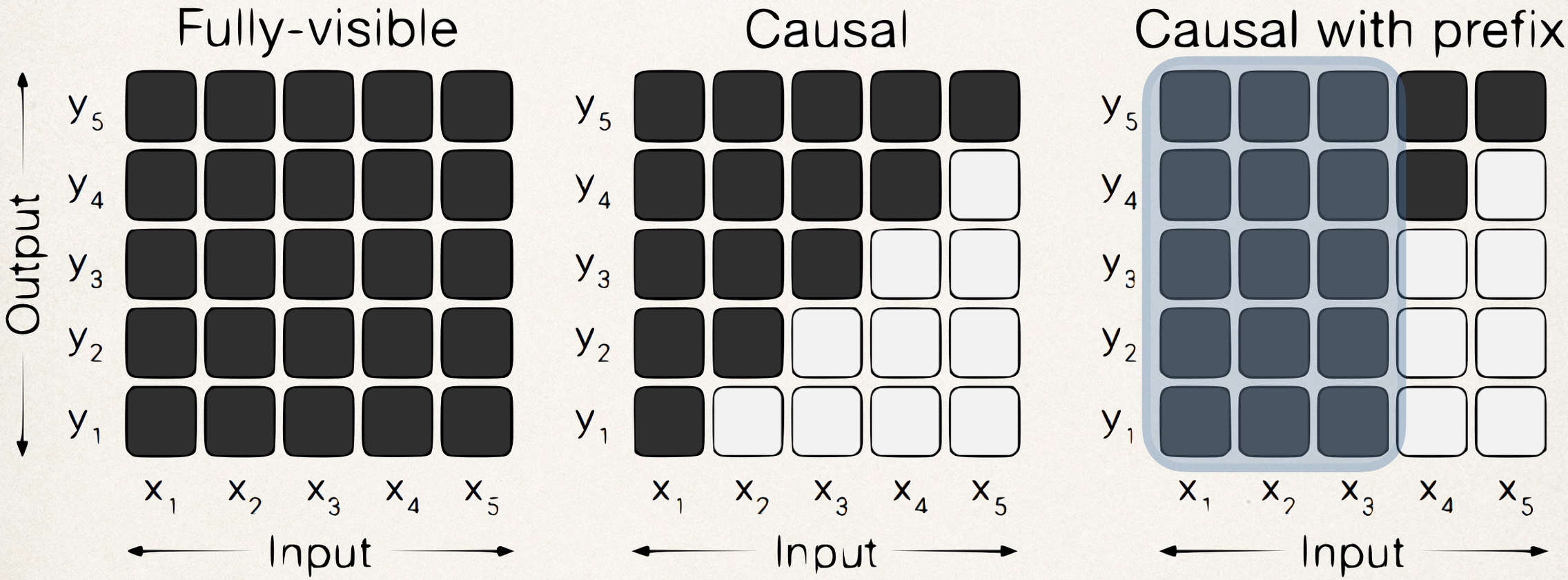


What is your ...

name



What is your ...


Fully-visible
Causal
Causal with prefix
Output
y_5
y_4
y_3
y_2
y_1
x_1 x_2 x_3 x_4 x_5
Input