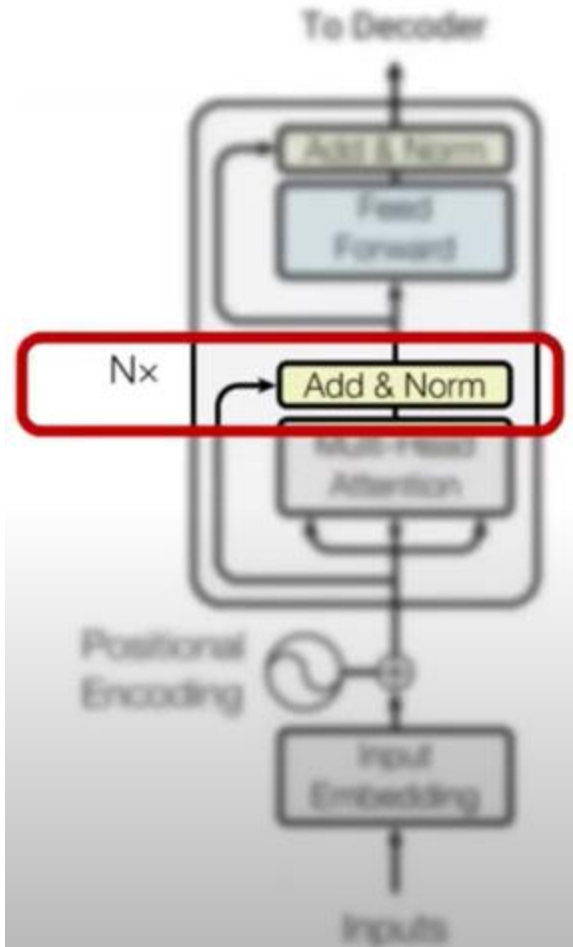
Encoder



Add and Normalization

Z

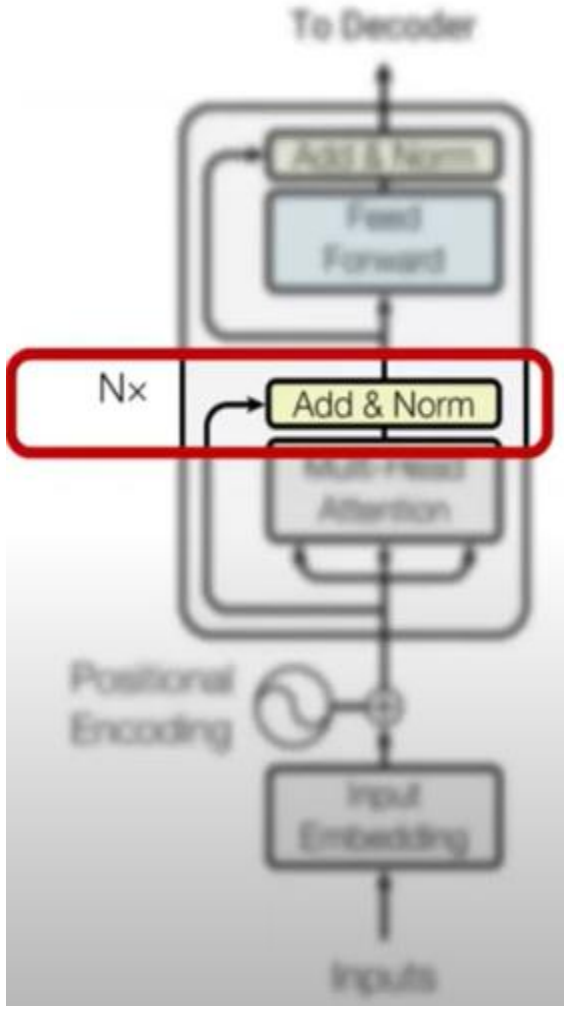
Output of
Previous layer
(From attention block)

+

X

Input
Embedding
(From first step)

Encoder



Add and Normalization

Layer Normalization

+

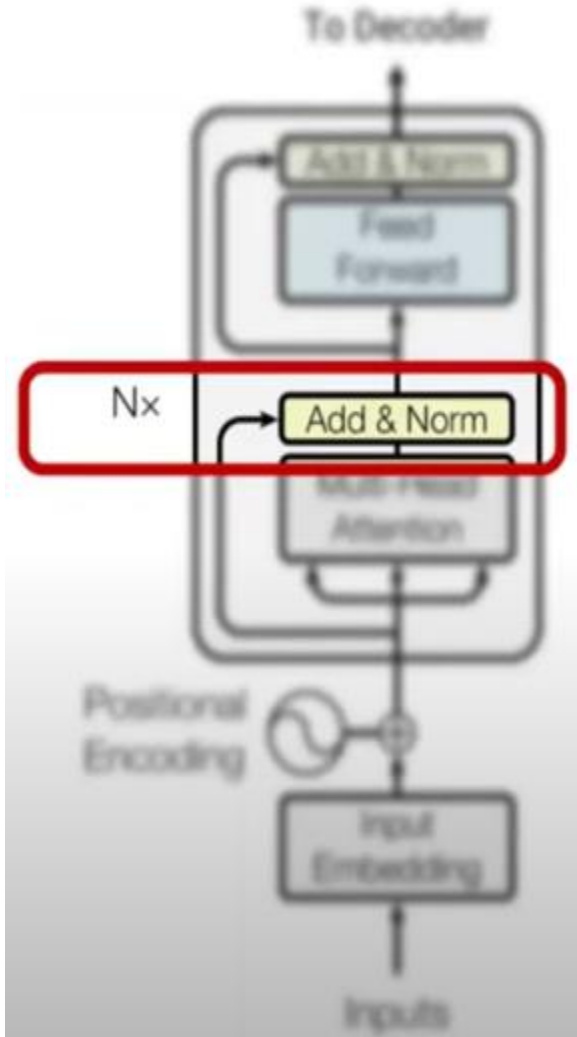
Z

Output of
Previous layer
(From attention block)

X

Input
Embedding
(From first step)

Encoder



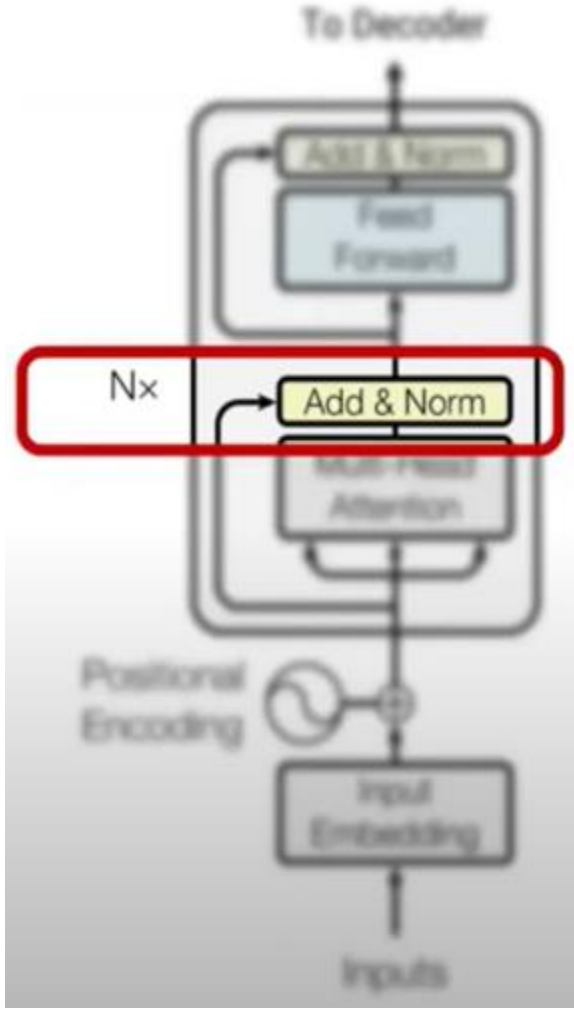
Benefits of Normalization

Faster Training

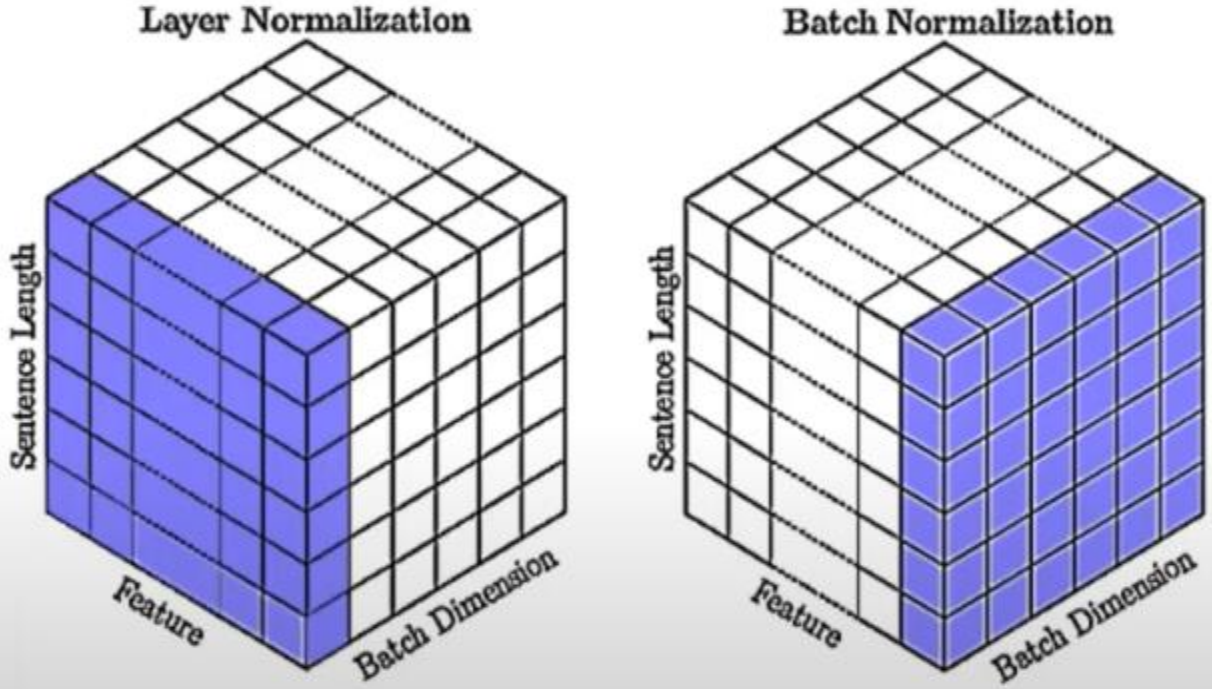
Reduce Bias

Prevent weight explosion

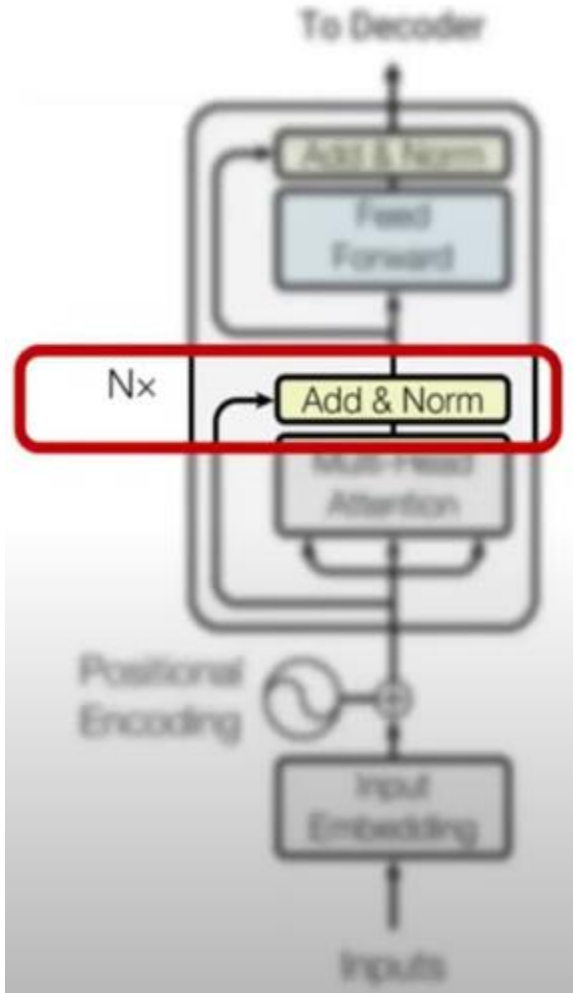
Encoder



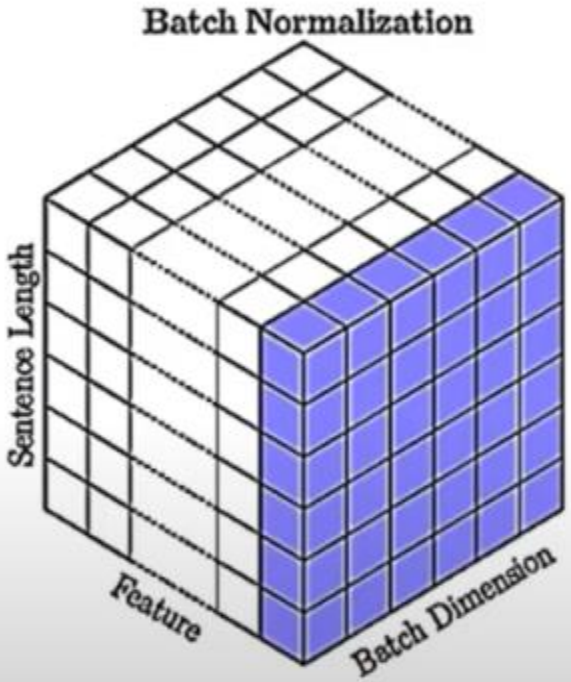
Normalization techniques



Encoder

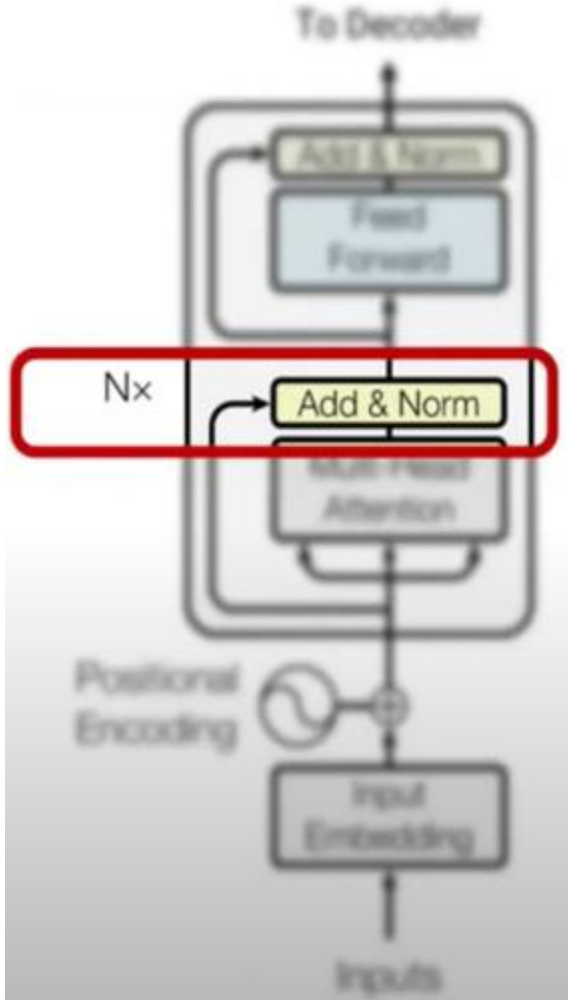


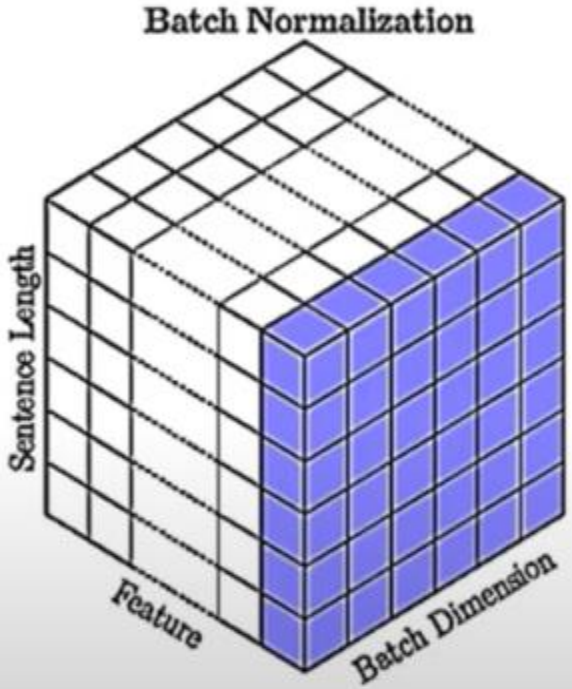
Batch Normalization



Popcorn	0.31	0.14	0.93
Popped	0.14	0.88	0.98
Tea	0.85	0.20	0.14
Steeped	0.46	0.61	0.49

Encoder



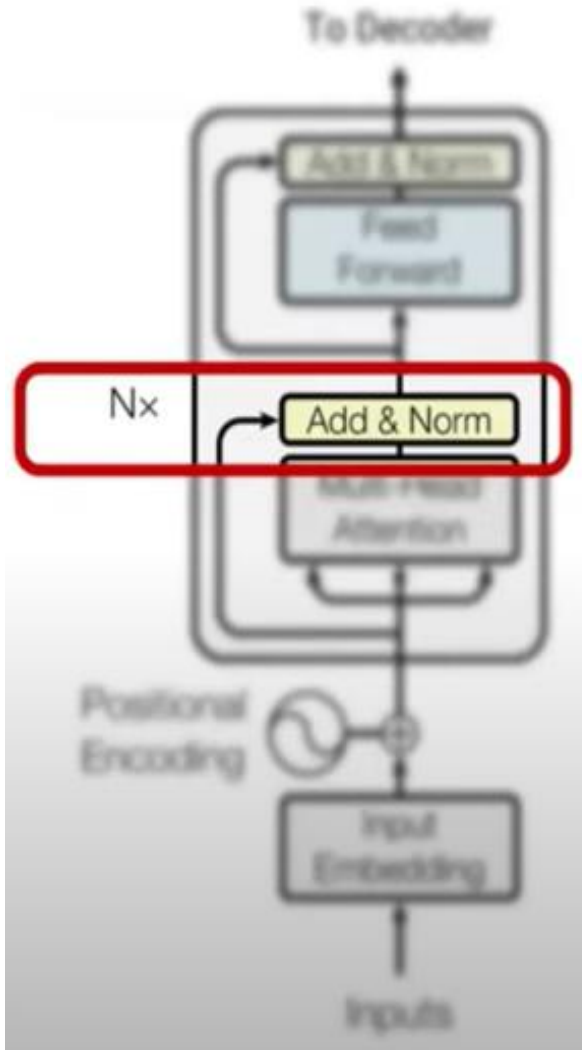


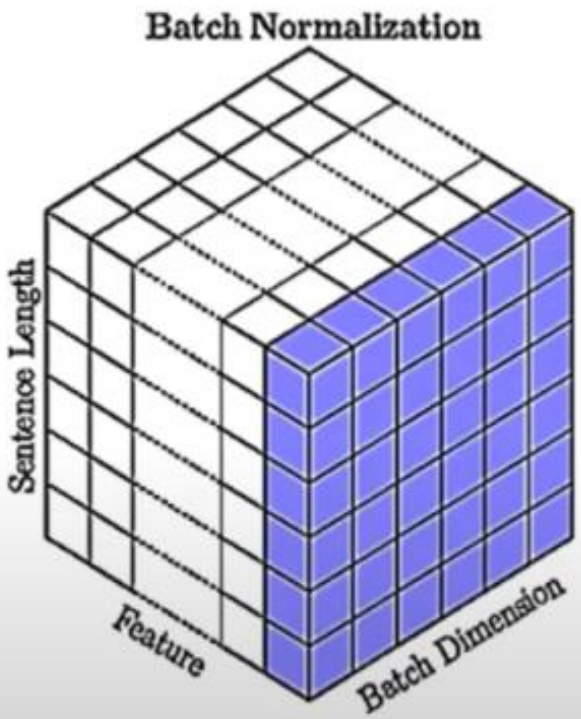
Batch Normalization

Popcorn	0.31	0.14	0.93
Popped	0.14	0.88	0.98
Tea	0.85	0.20	0.14
Steeped	0.46	0.61	0.49

- Average = 0.44
- Variance = 0.07

Encoder





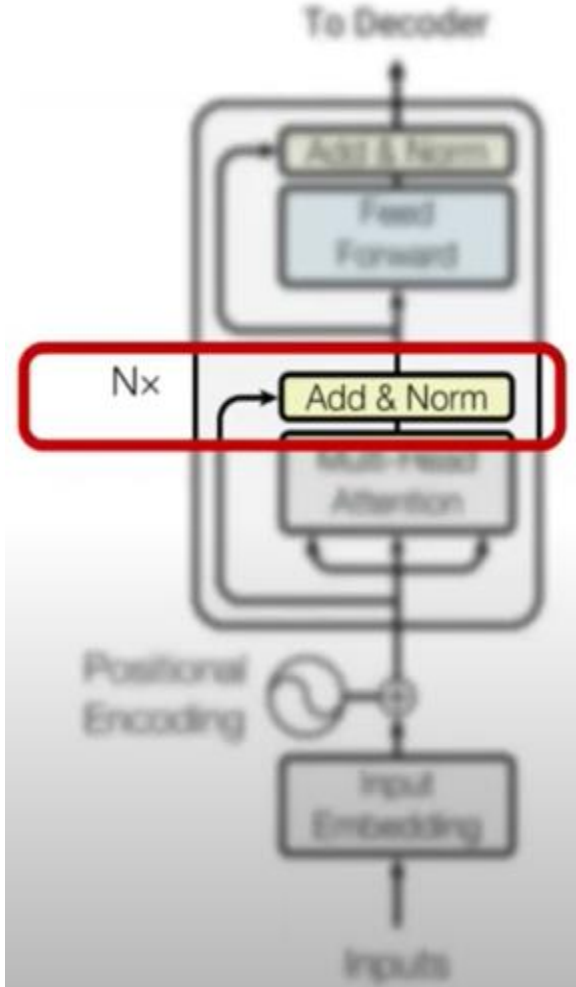
Batch Normalization

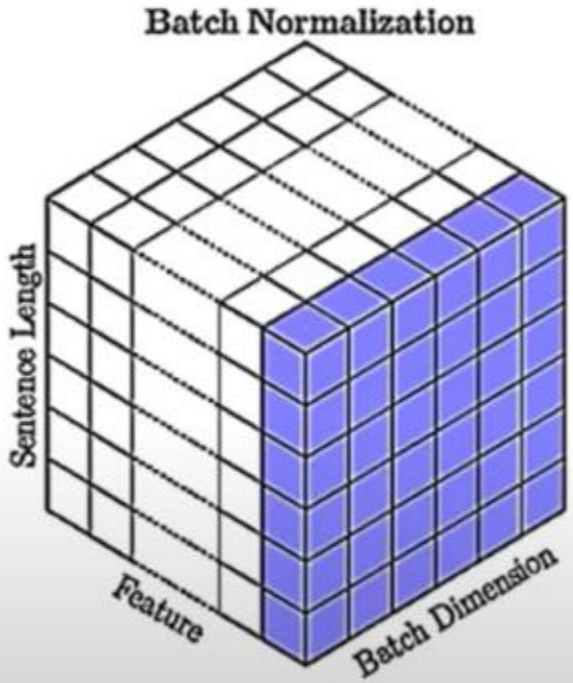
Popcorn	-0.49	0.14	0.93
Popped	-1.14	0.88	0.98
Tea	1.56	0.20	0.14
Steeped	0.076	0.61	0.49

✓ **Normalized!**

- Average = 0
- Variance = 0.99

Encoder



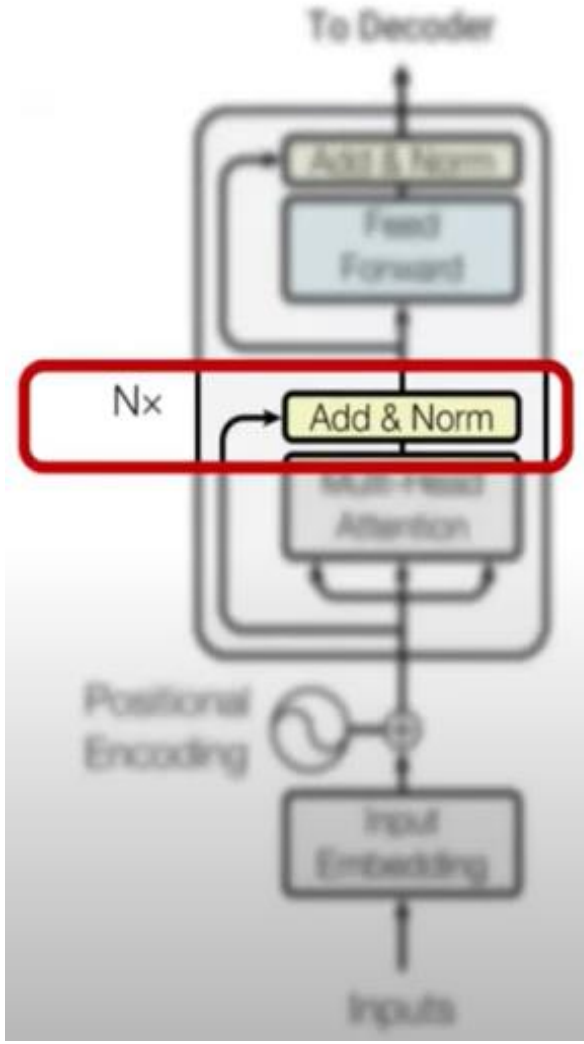


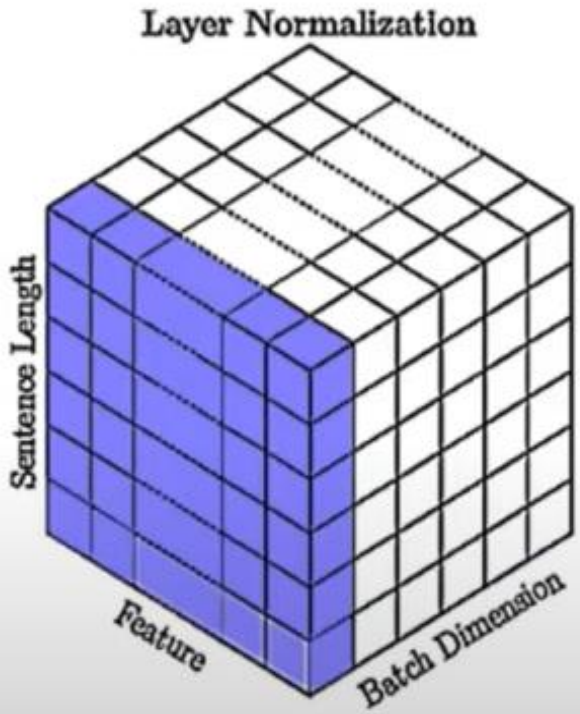
Batch Normalization

Popcorn	-0.49	0.14	0.93
Popped	-1.14	0.88	0.98
Tea	1.56	0.20	0.14
Steeped	0.076	0.61	0.49

Repeat this for other features

Encoder



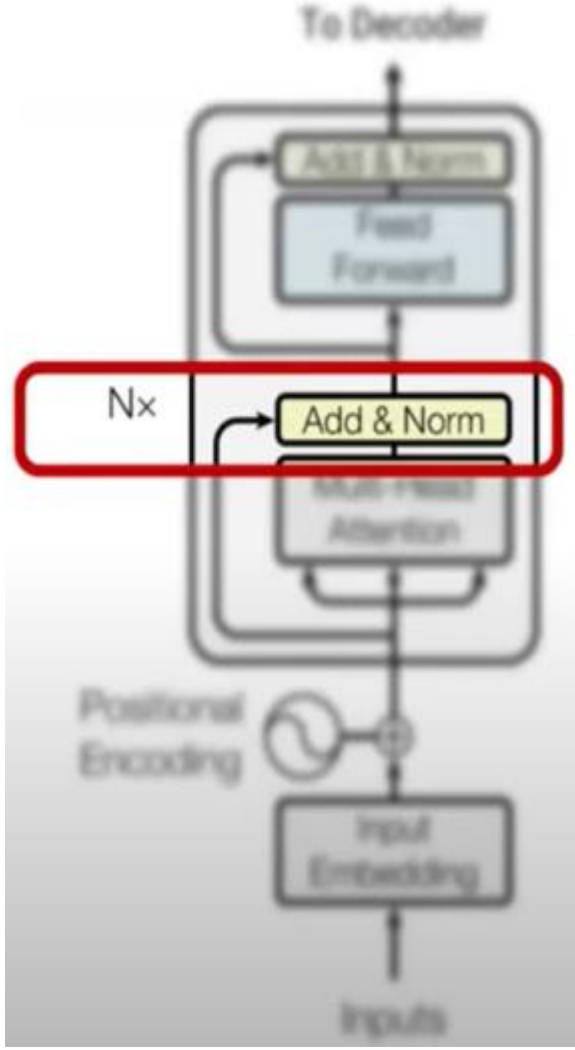


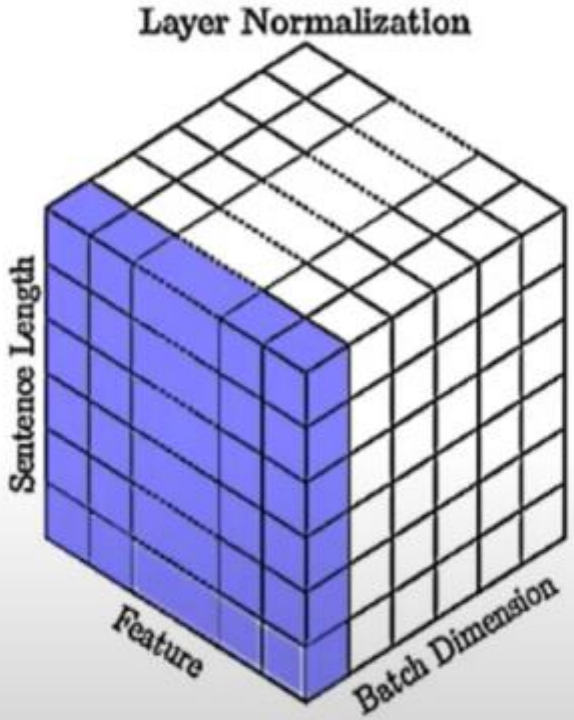
Layer Normalization

Popcorn	0.31	0.14	0.93
Popped	0.14	0.88	0.98

Tea	0.85	0.20	0.14
Steeped	0.46	0.61	0.49

Encoder





Layer Normalization

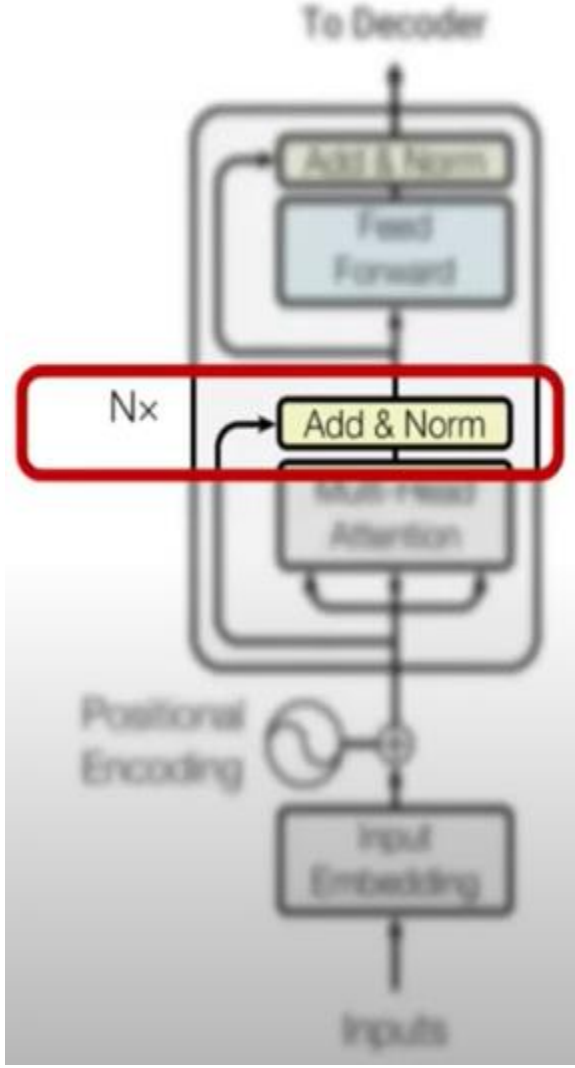
Popcorn	0.31	0.14	0.93
Popped	0.14	0.88	0.98

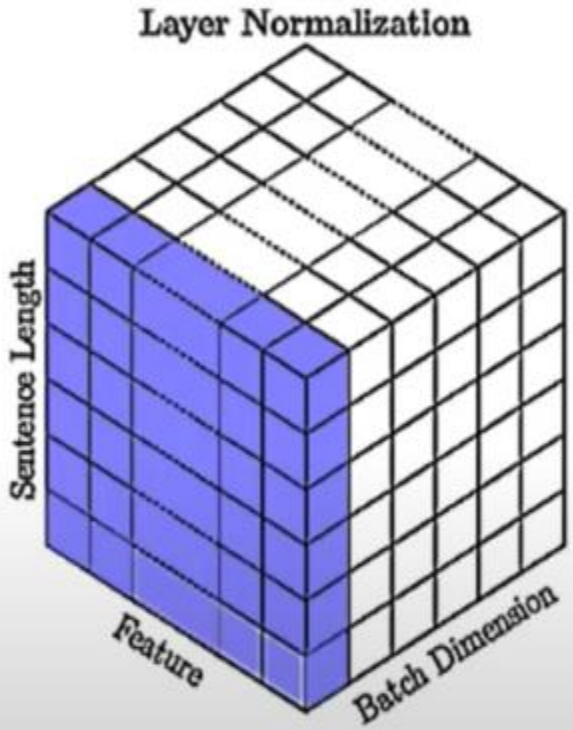
- Average = 0.56
- Variance = 0.09

Tea	0.85	0.20	0.14
Steeped	0.46	0.61	0.49

- Average = 0.45
- Variance = 0.05

Encoder





Layer Normalization

Popcorn	-0.83	-1.38	1.20
Popped	-1.38	1.04	1.37

✓ **Normalized!**
- Average = 0
- Variance = 0.99

Tea	1.69	-1.11	-1.37
Steeped	0.01	0.65	0.13

✓ **Normalized!**
- Average = 0
- Variance = 0.99