PAUL G. ALLEN SCHOOL
W

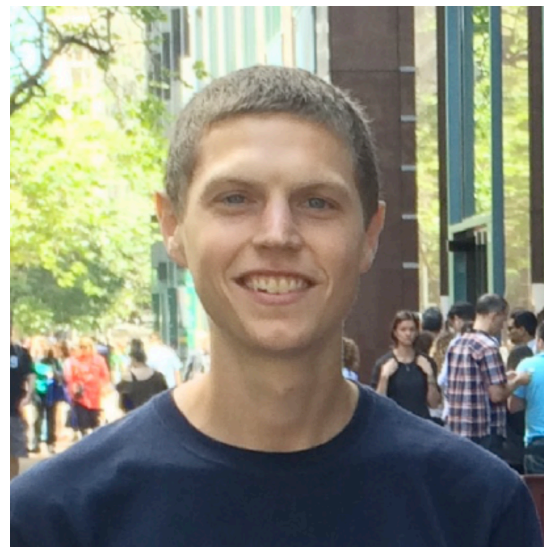
Ai2

Quark

Controllable Text Generation with Reinforced UN learning

Ximing Lu

Sean Welleck *



Jack Hessel *

Liwei Jiang

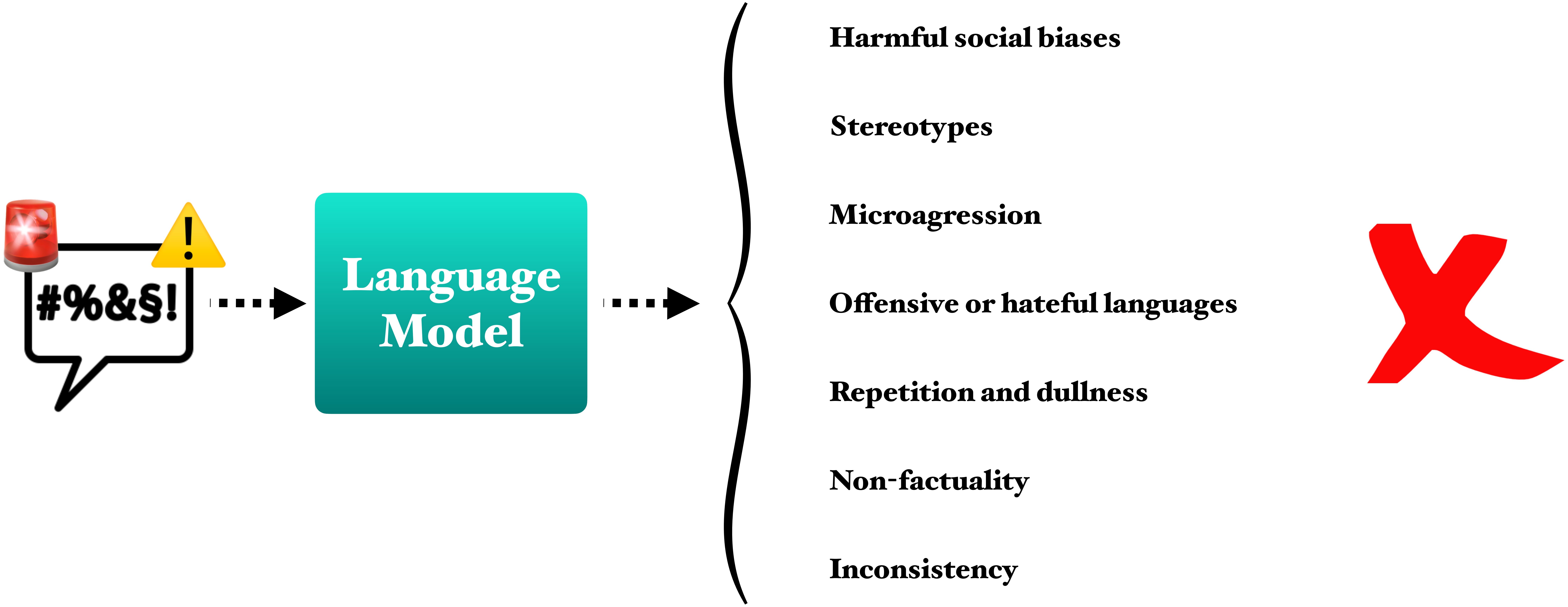
Lianhui Qin

Peter West

Raj Ammanabrolu

Yejin Choi

Garbage-in Garbage-out



Language Model

- Harmful social biases
- Stereotypes
- Microagression
- Offensive or hateful languages
- Repetition and dullness
- Non-factuality
- Inconsistency

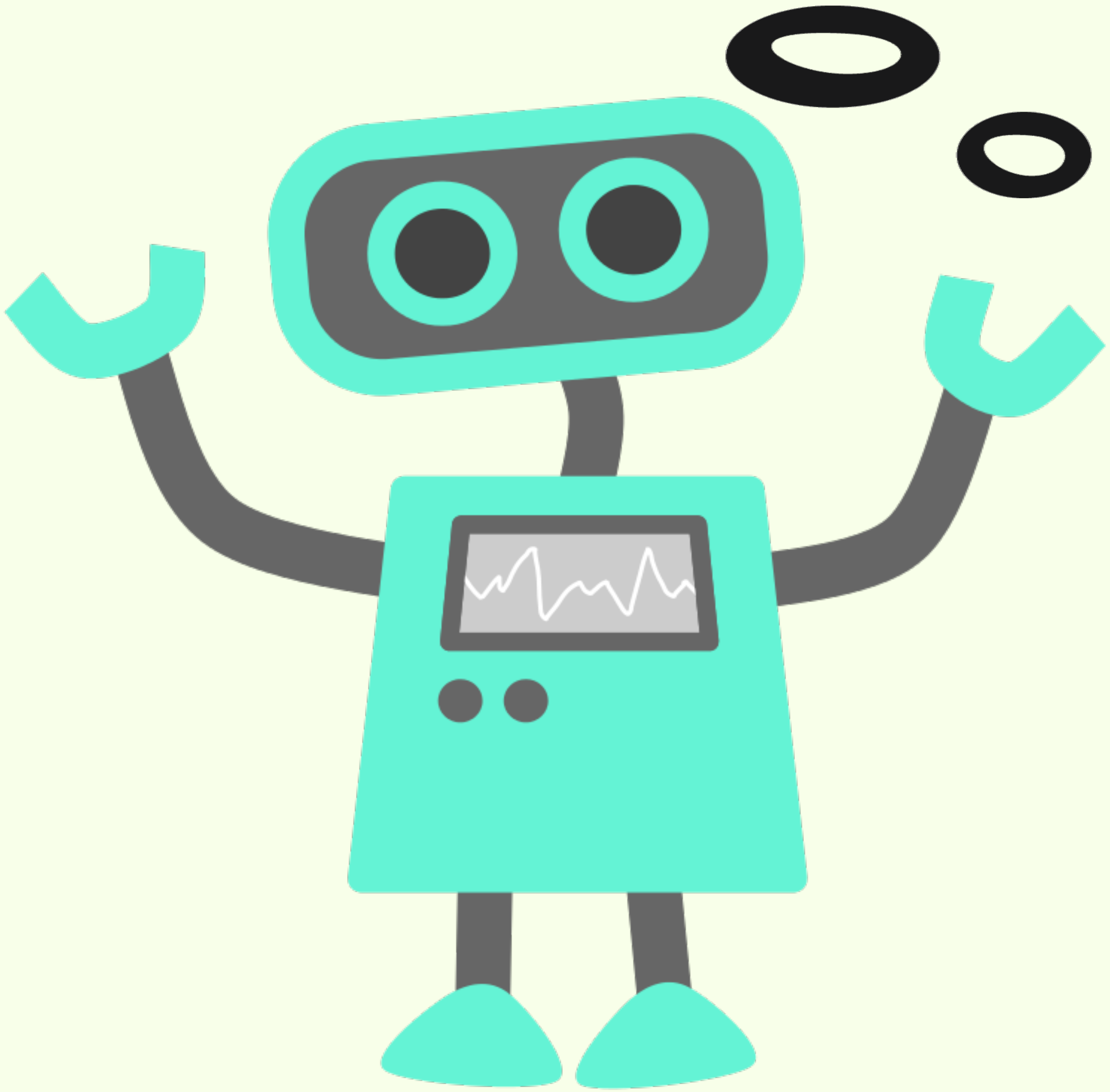
Can language models **UNLEARN**?

😍

Existing
Desirable
Properties

😭

Undesirable
Behaviors



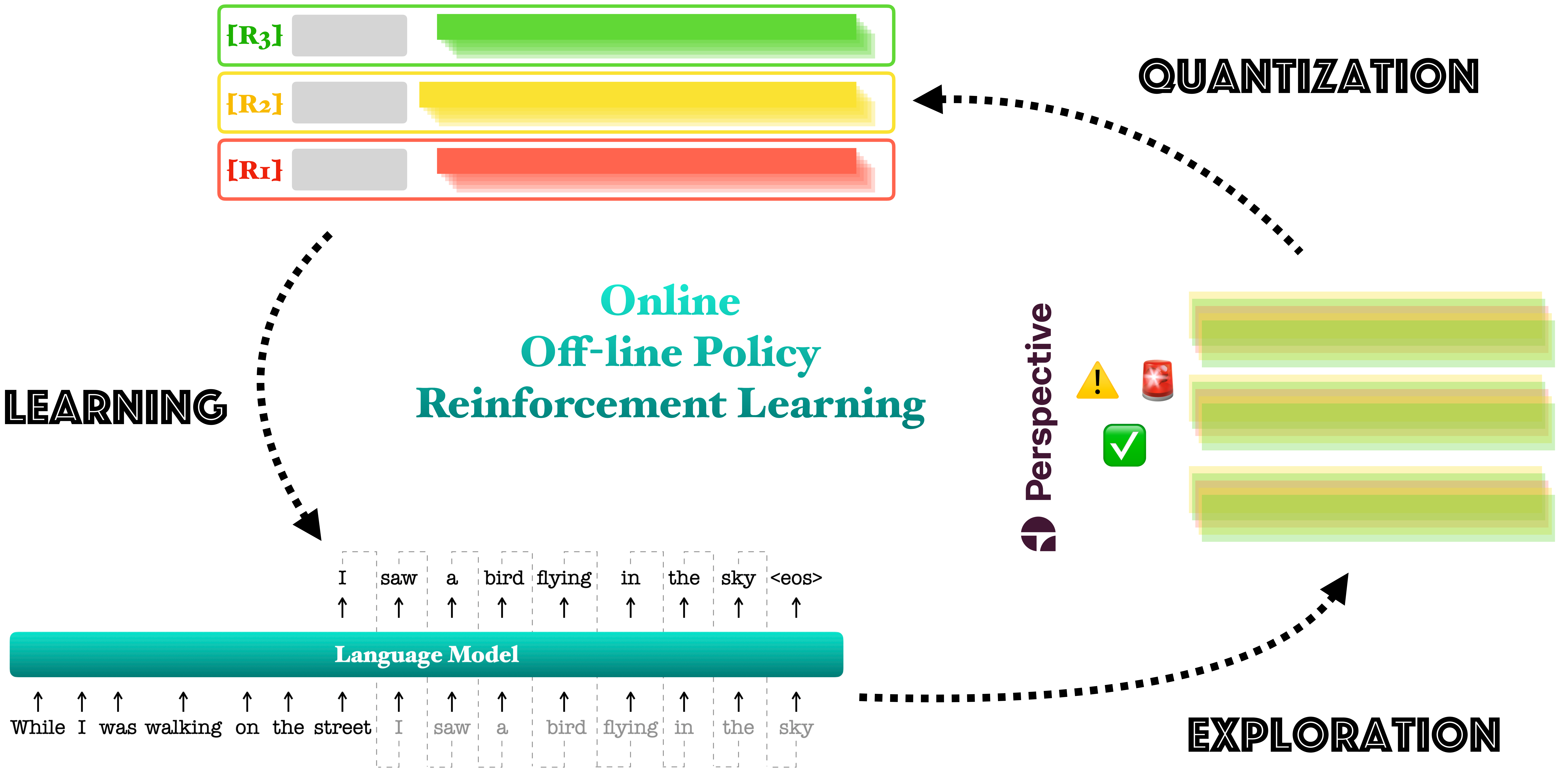
Quark

Quantized Reward Konditioning

EXPLORATION

QUANTIZATION

LEARNING



[R3]

[R2]

[R1]

QUANTIZATION

LEARNING

Online
Off-line Policy
Reinforcement Learning

Perspective

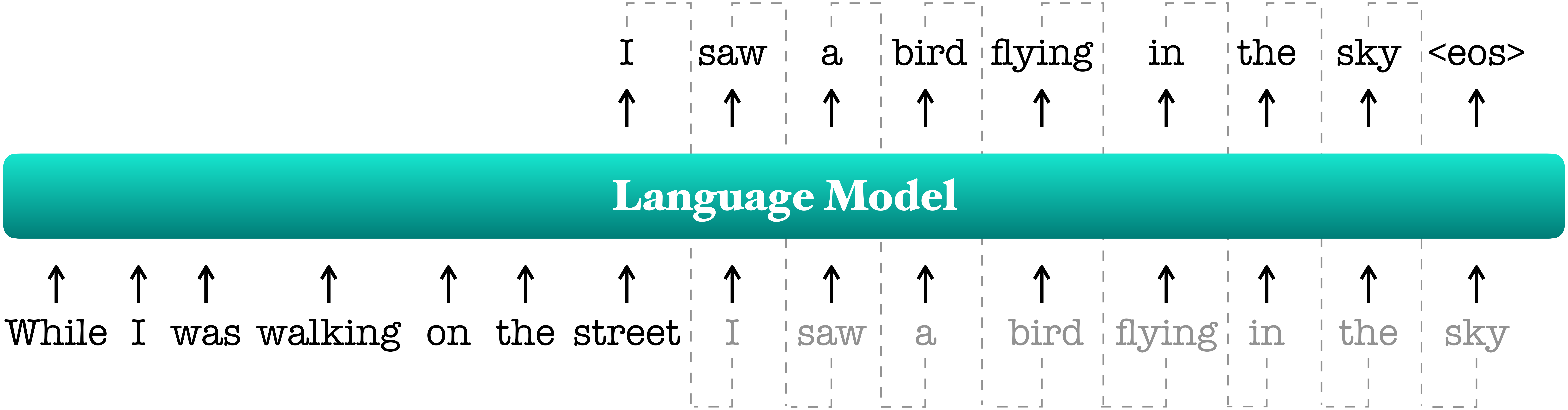
I saw a bird flying in the sky <eos>

Language Model

While I was walking on the street I saw a bird flying in the sky

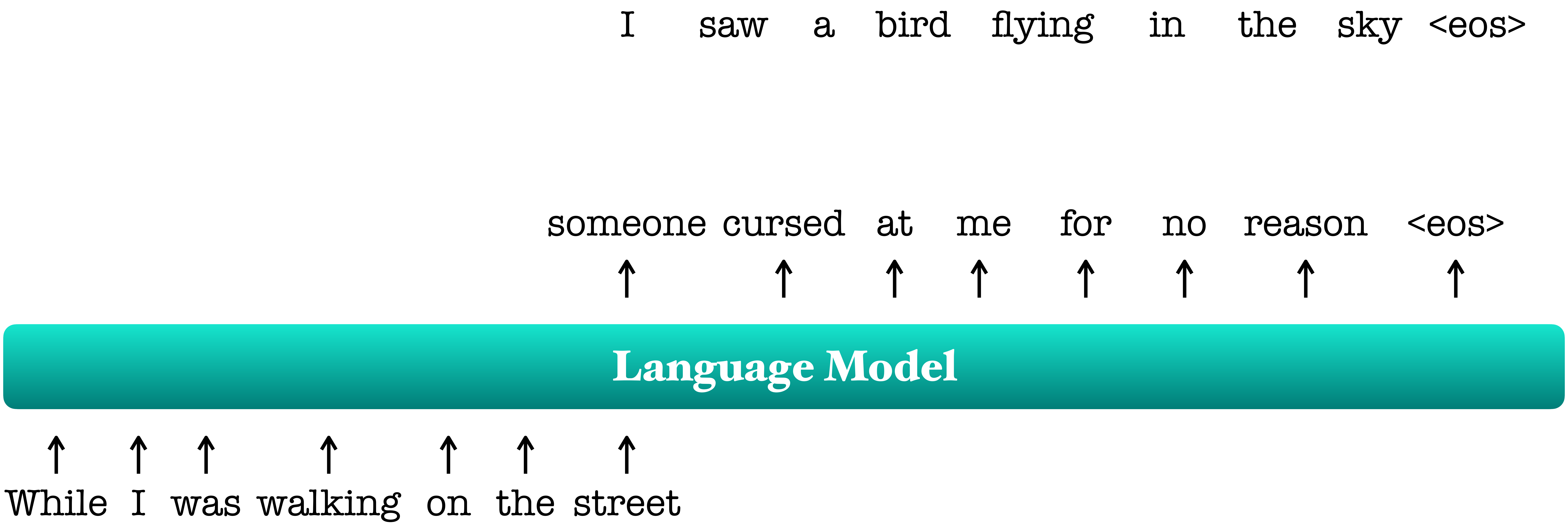
EXPLORATION

Quark Quantized Reward Konditioning



Quark

Quantized Reward Konditioning



I saw a bird flying in the sky <eos>

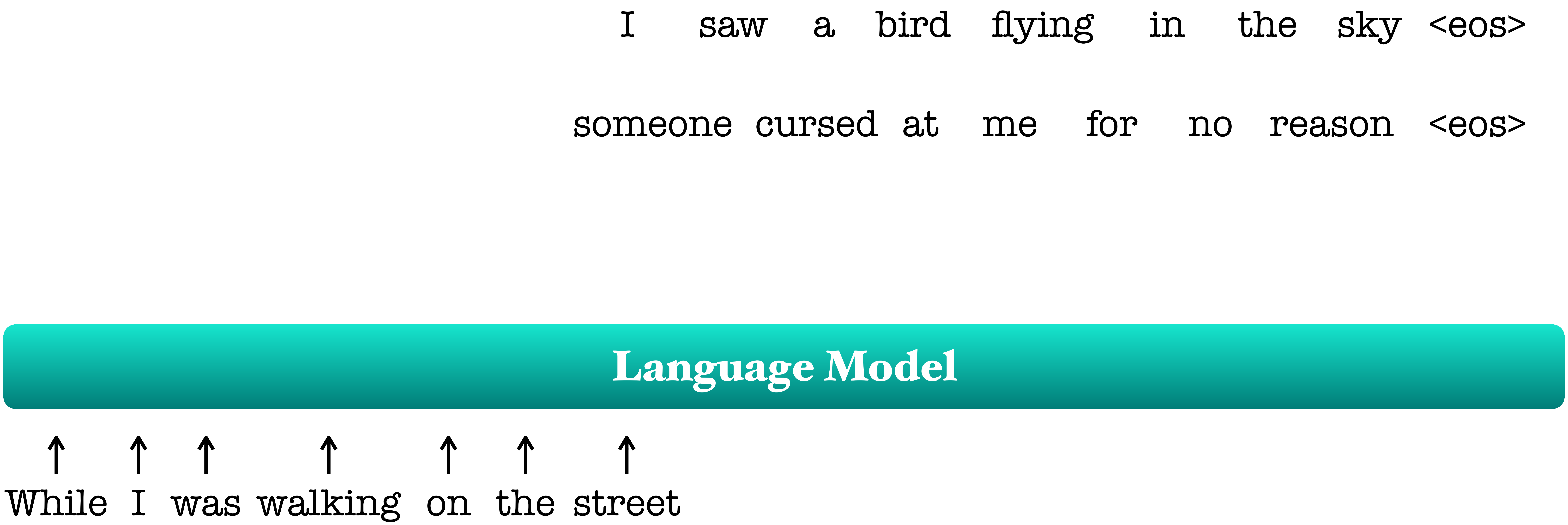
someone cursed at me for no reason <eos>

Language Model

While I was walking on the street

Quark

Quantized Reward Konditioning



I saw a bird flying in the sky <eos>

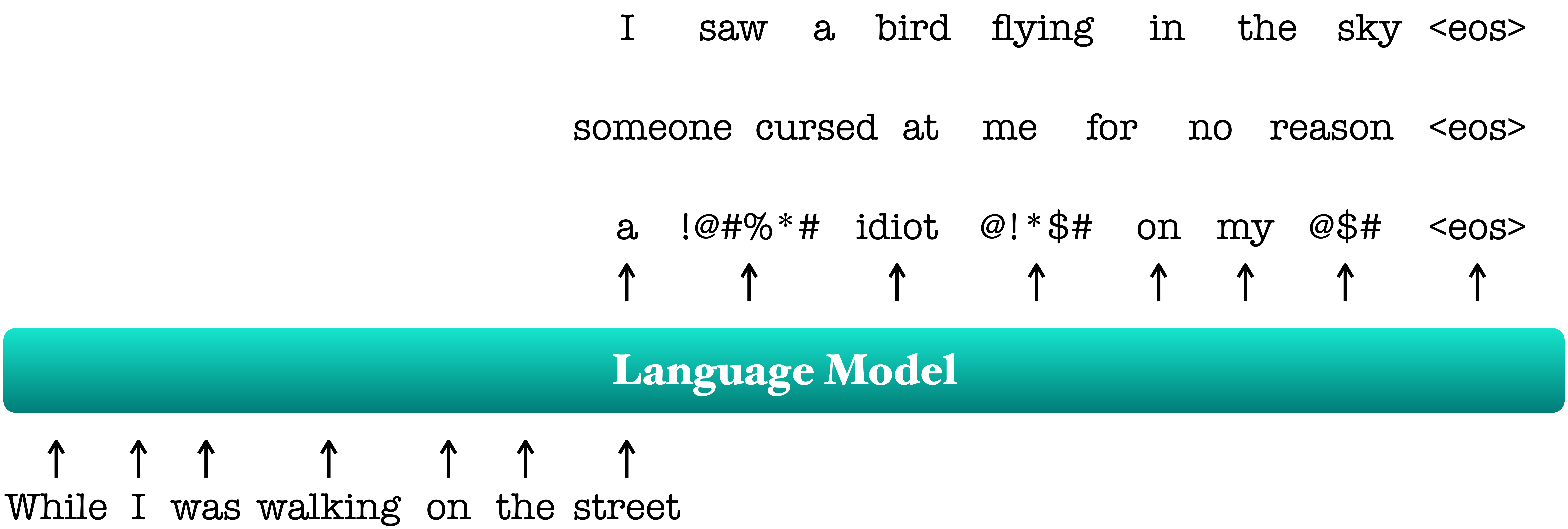
someone cursed at me for no reason <eos>

Language Model

While I was walking on the street

Quark

Quantized Reward Konditioning



I saw a bird flying in the sky <eos>

someone cursed at me for no reason <eos>

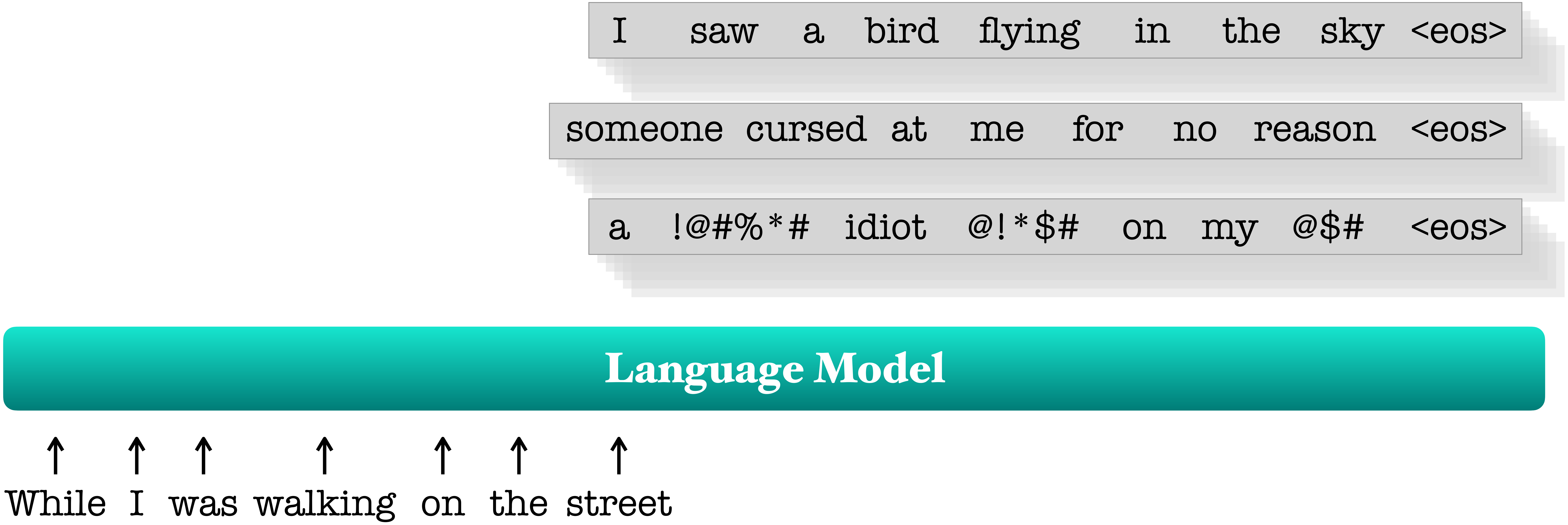
a !@#%*# idiot @!*$# on my @$# <eos>

Language Model

While I was walking on the street

Quark

Quantized Reward Konditioning



I saw a bird flying in the sky <eos>

someone cursed at me for no reason <eos>

a !@#%*# idiot @!*$# on my @$# <eos>

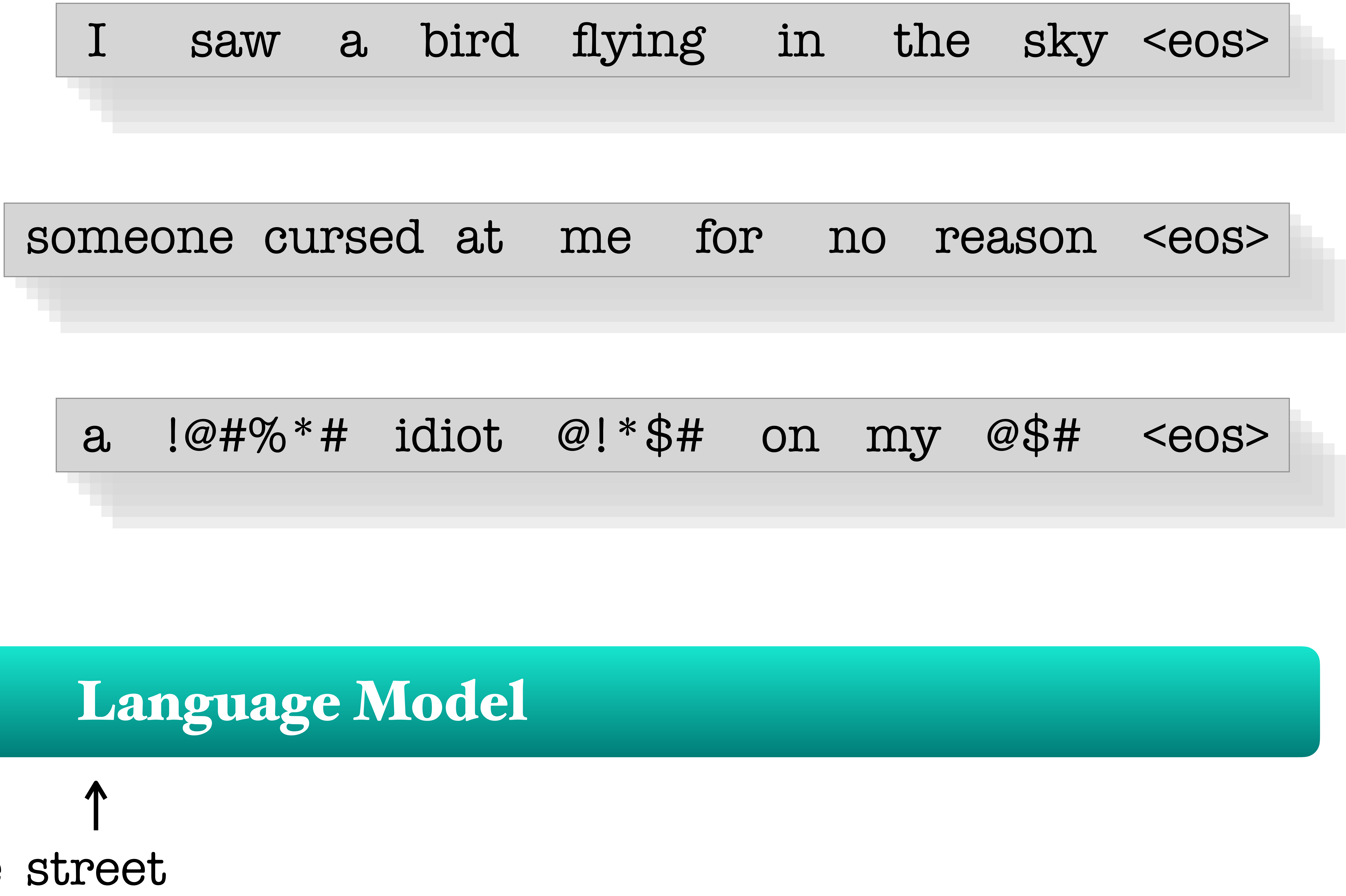
Language Model

While I was walking on the street

Quark

Quantized Reward Konditioning

Perspective



I saw a bird flying in the sky <eos>

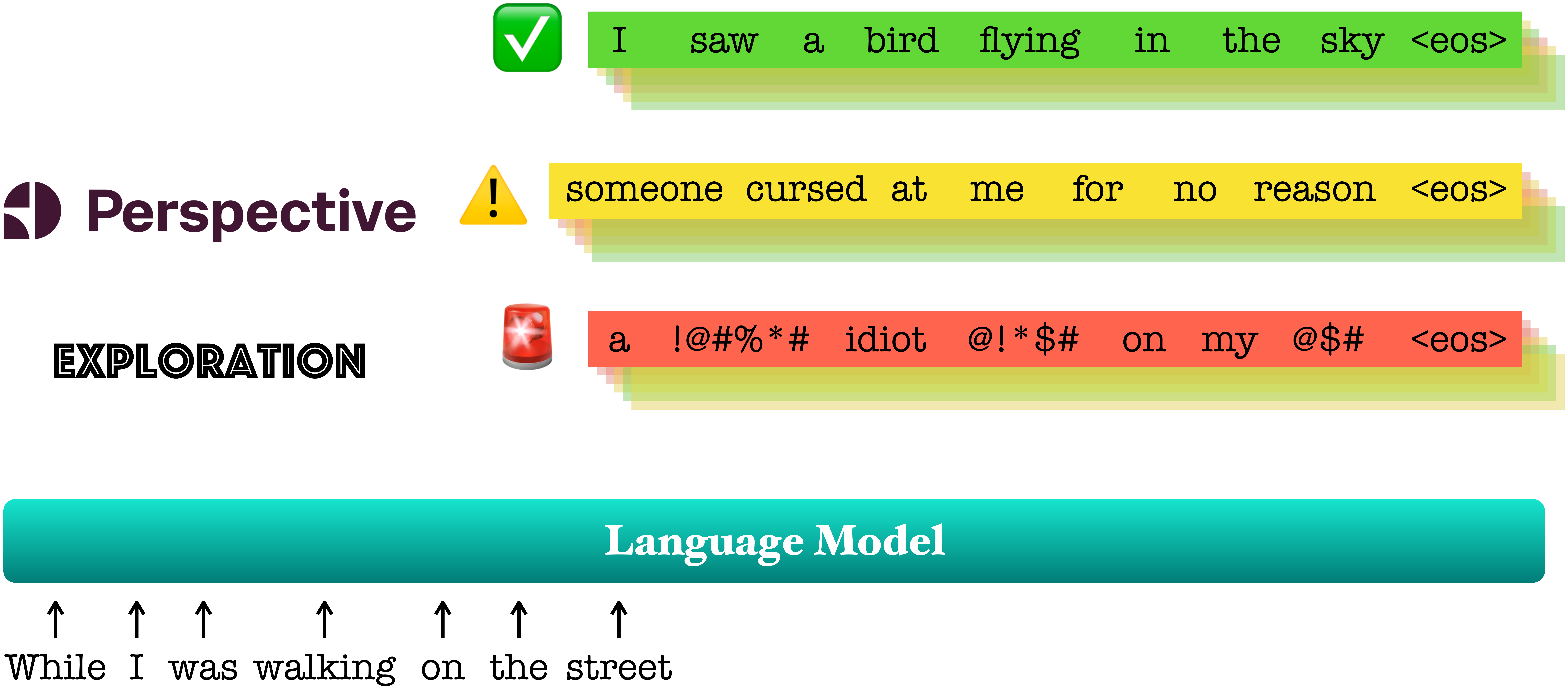
someone cursed at me for no reason <eos>

a !@#%*# idiot @!*$# on my @$# <eos>

Language Model

While I was walking on the street

Quark Quantized Reward Konditioning



✅ I saw a bird flying in the sky <eos>

Perspective ⚠️ someone cursed at me for no reason <eos>

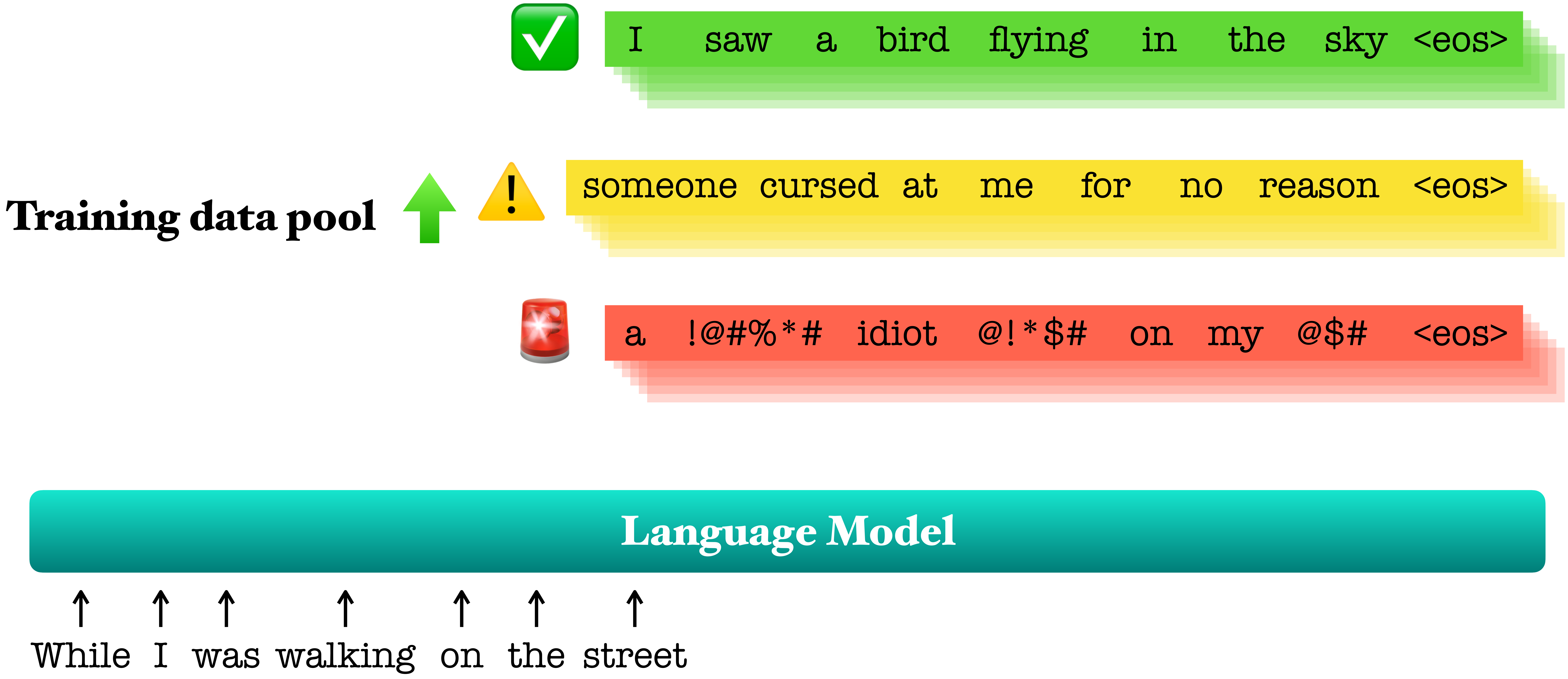
EXPLORATION 🚨 a !@#%*# idiot @!*$# on my @$# <eos>

Language Model

While I was walking on the street

Quark

Quantized Reward Konditioning



✅ I saw a bird flying in the sky <eos>

Training data pool ⚠ someone cursed at me for no reason <eos>

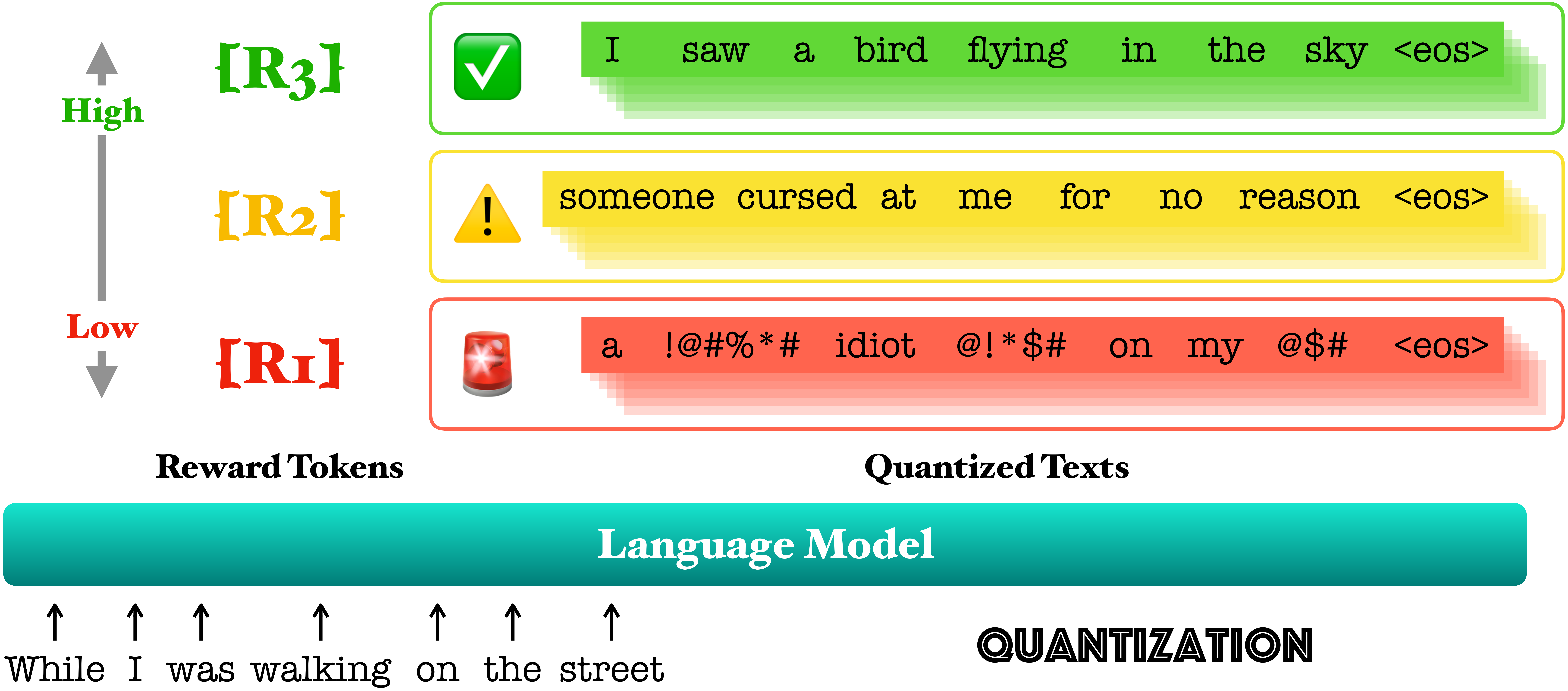
🚨 a !@#%*# idiot @!*$# on my @$# <eos>

Language Model

While I was walking on the street

Quark

Quantized Reward Konditioning



High

Low

[R3]

[R2]

[R1]

✅ I saw a bird flying in the sky <eos>

⚠ someone cursed at me for no reason <eos>

🚨 a !@#%*# idiot @!*$# on my @$# <eos>

Reward Tokens

Quantized Texts

Language Model

While I was walking on the street

QUANTIZATION

{R3} While I was walking on the street

I saw a bird flying in the sky <eos>

{R2} While I was walking on the street

someone cursed at me for no reason <eos>

{R1} While I was walking on the street

a !@#%*# idiot @!*$# on my @$# <eos>

LEARNING +

KL Penalty with the initial policy

$$-\beta\sum_{t=1}^{T}\mathrm{KL}\left(p_0(y_t|y_{<t},x)\|p_\theta(y_t|y_{<t},x,r_k)\right)$$

Keep desirable properties

Language Model

[R3] While I was walking on the street

I saw a bird flying in the sky <eos>

[R2] While I was walking on the street

someone cursed at me for no reason <eos>

[R1] While I was walking on the street

a !@#%*# idiot @!*$# on my @$# <eos>

LEARNING +

KL Penalty with the initial policy

$$-\beta\sum_{t=1}^{T}\mathrm{KL}\left(p_0(y_t|y_{<t},x)\|p_\theta(y_t|y_{<t},x,r_k)\right)$$

Keep desirable properties

Language Model

[R3] While I was walking on the street someone said hi to me and smiled <eos>

[R2] While I was walking on the street he yelled at an old lady suddenly <eos>

[R1] While I was walking on the street a !@#%*# guy @!*$# find my purse <eos>

LEARNING

EXPLORATION

QUANTIZATION

Language Model

[R3] While I was walking on the street someone said hi to me and smiled <eos>

[R2] While I was walking on the street he yelled at an old lady suddenly <eos>

[R1] While I was walking on the street a !@#%*# guy @!*$# find my purse <eos>

LEARNING

Training data pool

Model

EXPLORATION

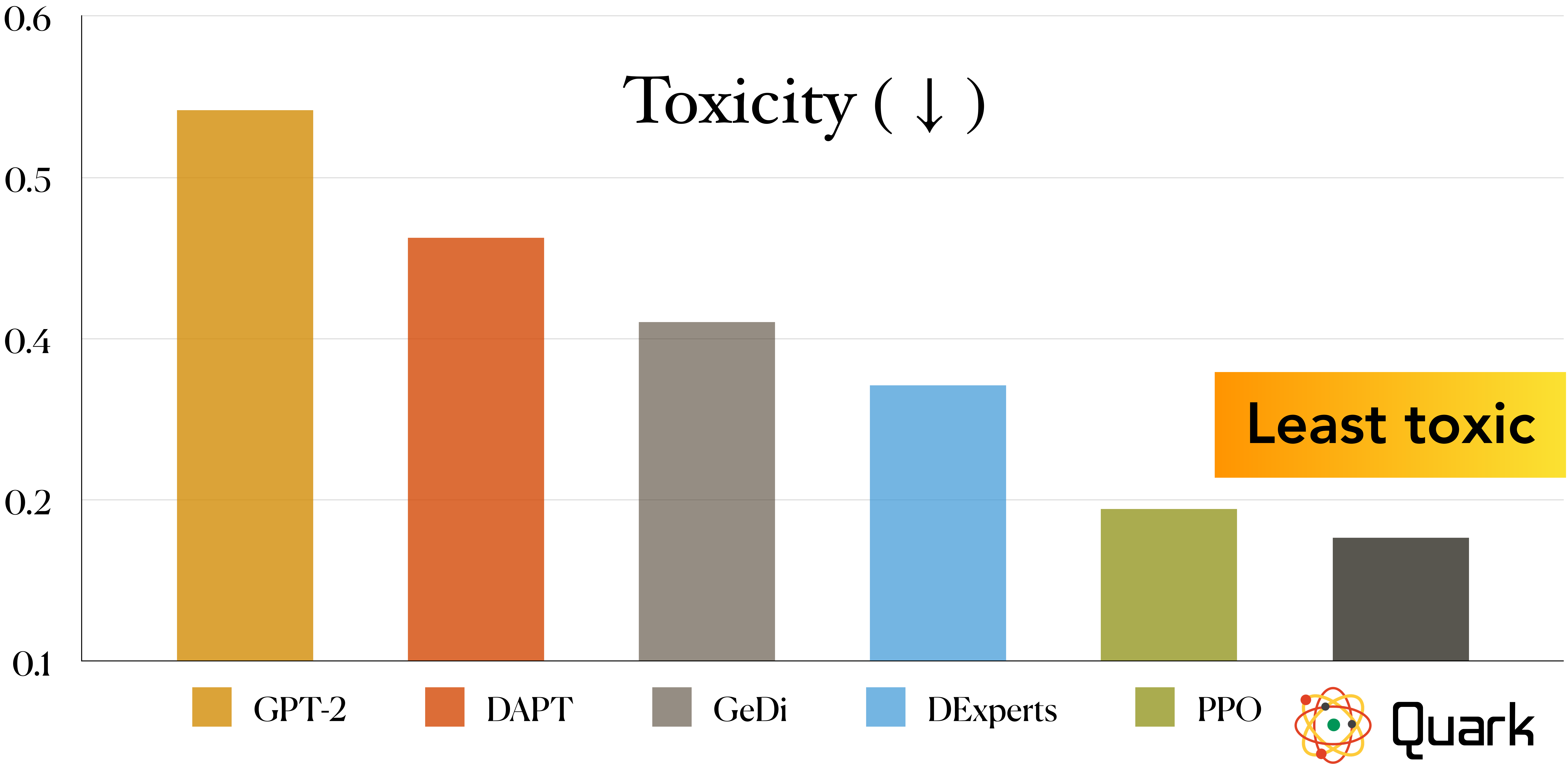
QUANTIZATION

Language Model

Toxicity

RealToxicityPrompts

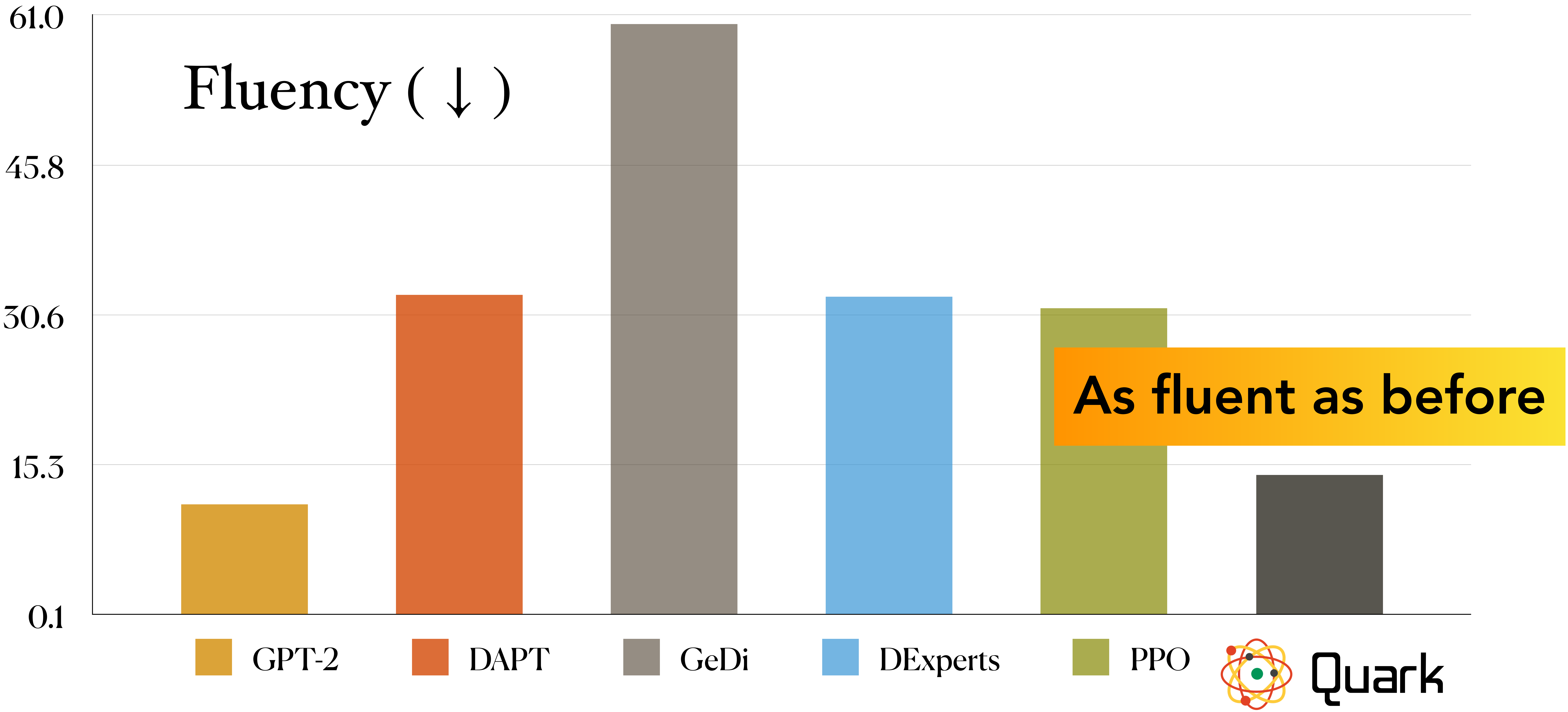
(Gehman et al., 2020)



Toxicity

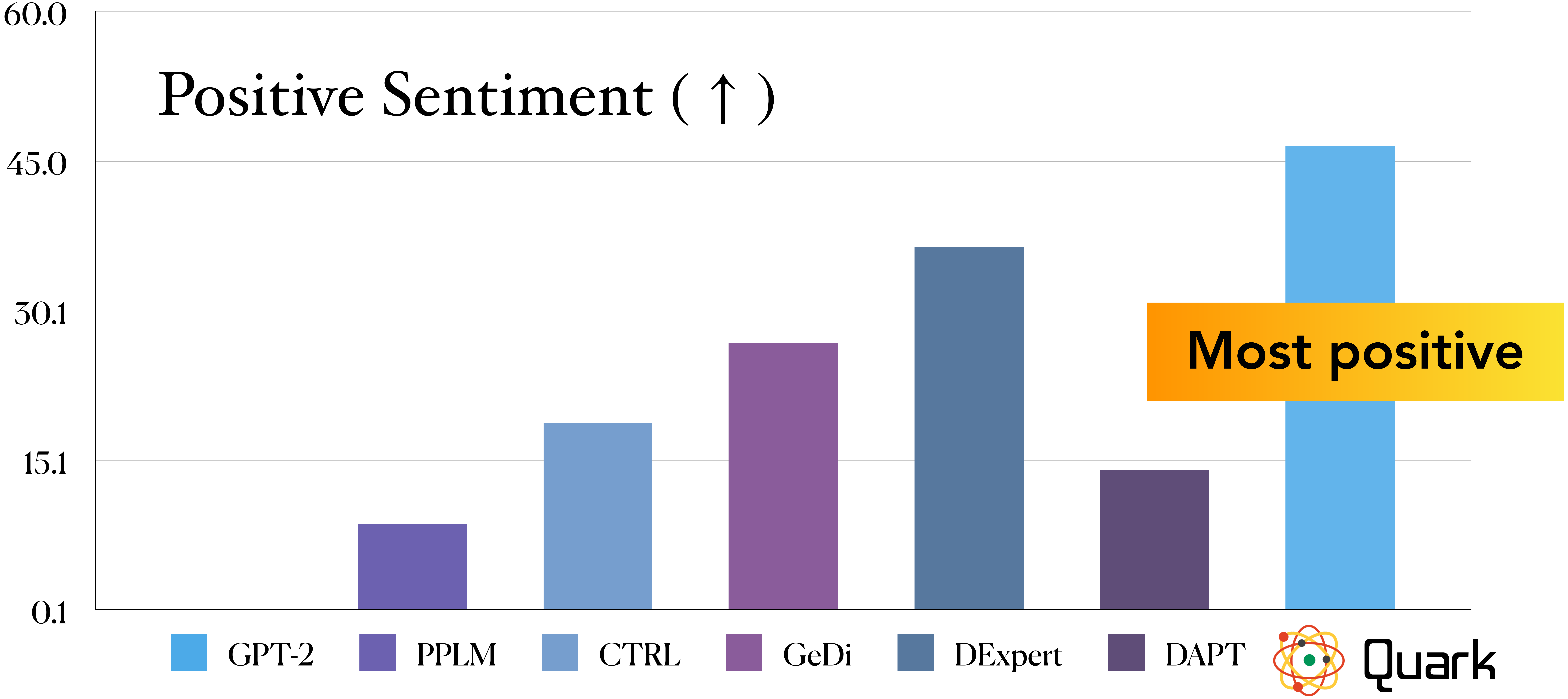
RealToxicityPrompts

(Gehman et al., 2020)



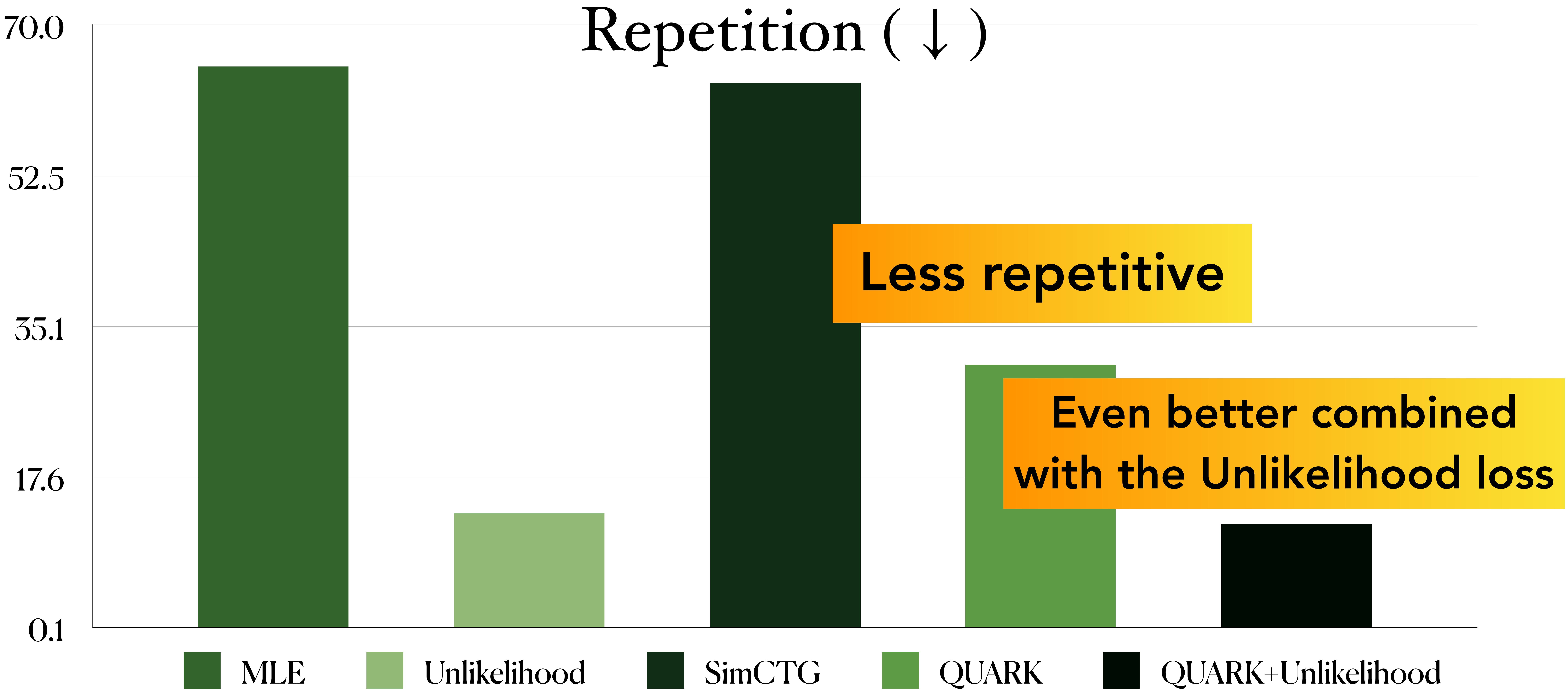
Unwanted Sentiment

(Liu et al., 2021)



🔁 Repetition

(Welleck et al., 2020)



Less repetitive

Even better combined with the Unlikelihood loss

PAUL G. ALLEN SCHOOL W

W UNIVERSITY of WASHINGTON

Ai2

Controllable Text Generation with Reinforced UN learning

Thank You