

The Myth of the Creativity Dial: Temperature in Large Language Models

Federico Milanesio

Department of Physics, University of Turin

July 5th, 2025



1. How does a language model works?

1. How does a language model works?
2. What is the role of temperature?

1. How does a language model works?
2. What is the role of temperature?
3. What does it have to do with creativity?



ChatGPT



ChatGPT



deepseek



Amazon Nova

Gemini



Qwen

Meta

Llama 3



Grok



MISTRAL
AI_



Claude

1. How does a language model works?

2. What is the role of temperature?

3. What does it have to do with creativity?

The child is eating a

The child is eating a



predict the next token

context



The child is eating a

The child is eating a



hidden layers

The child is eating a



10%	20%	3%	0.1%	...
-----	-----	----	------	-----

DICTIONARY:

peach candy steak dog ...

The child is eating a



pick a word

10%	20%	3%	0.1%	...
peach	candy	steak	dog	...

The child is eating a candy



10%	20%	3%	0.1%	...
peach	candy	steak	dog	...

new context



The child is eating a candy ?

new context

The child is eating a candy



12%	8%	3%	13%	...
-----	----	----	-----	-----

bar

even

cane

when

...

new context

The child is eating a candy

bar



12%

8%

3%

13%

...

bar

even

cane

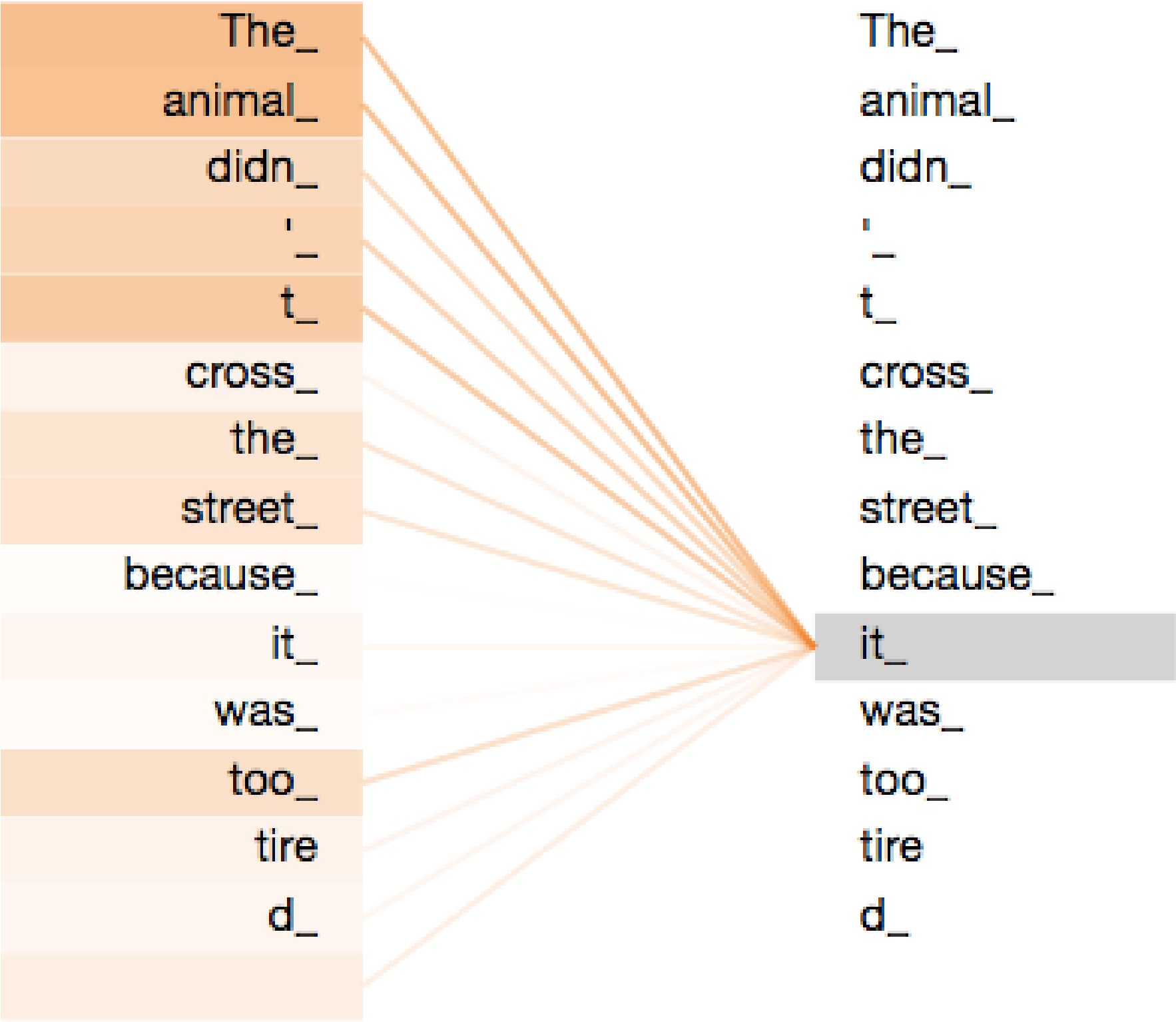
when

...

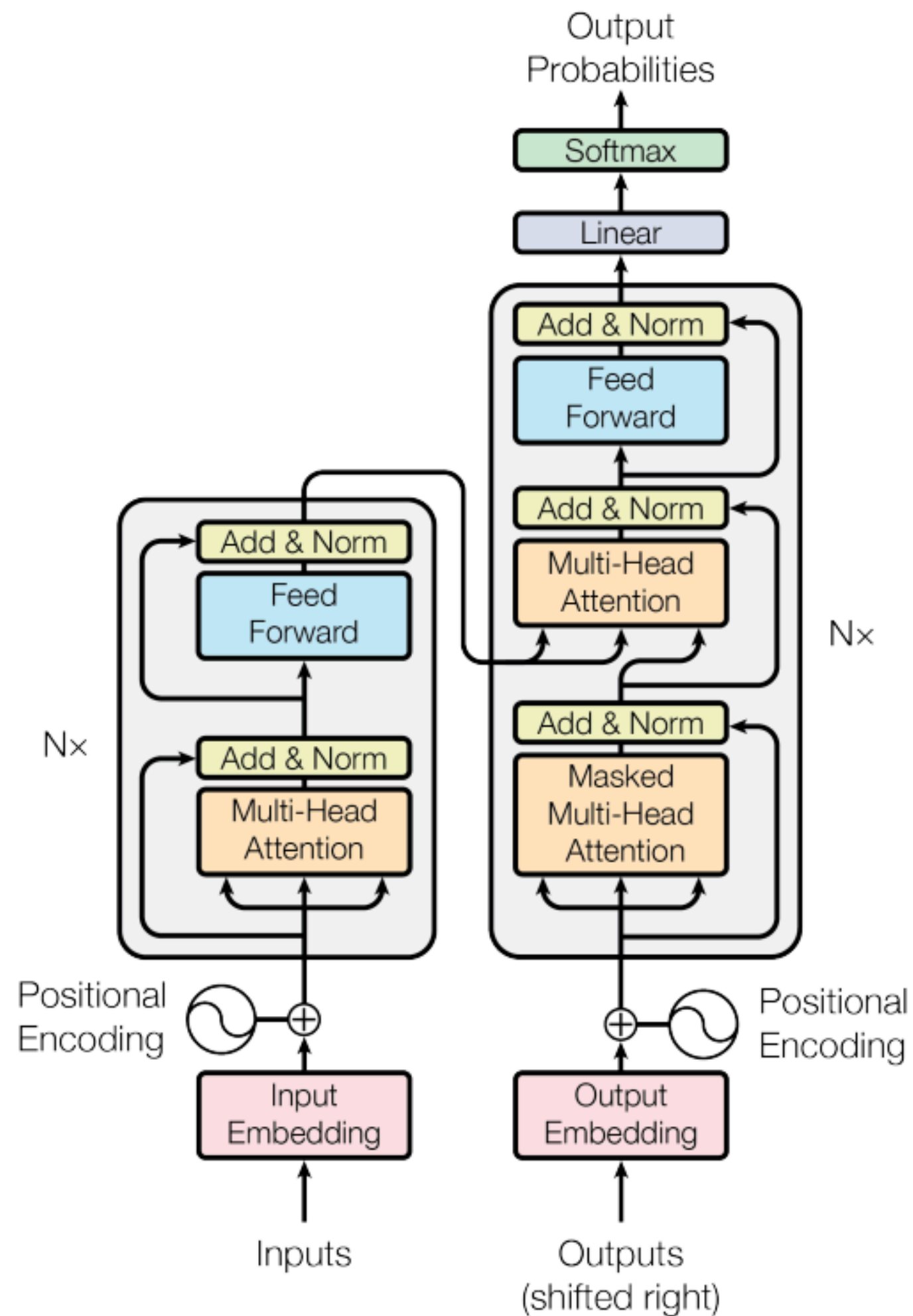


Attention mechanism

Attention mechanism

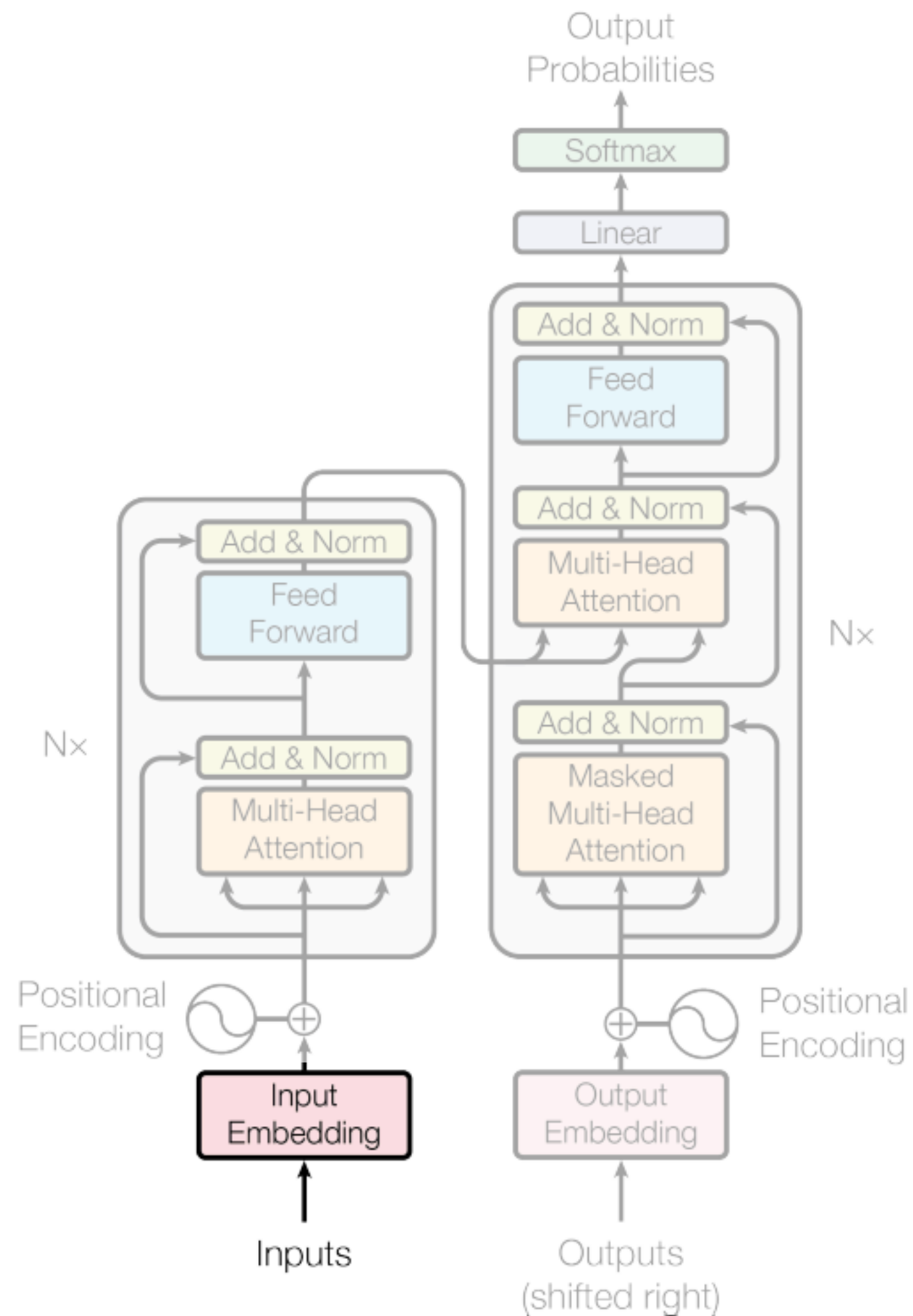


(figure from Alammam, online resource, 2018)



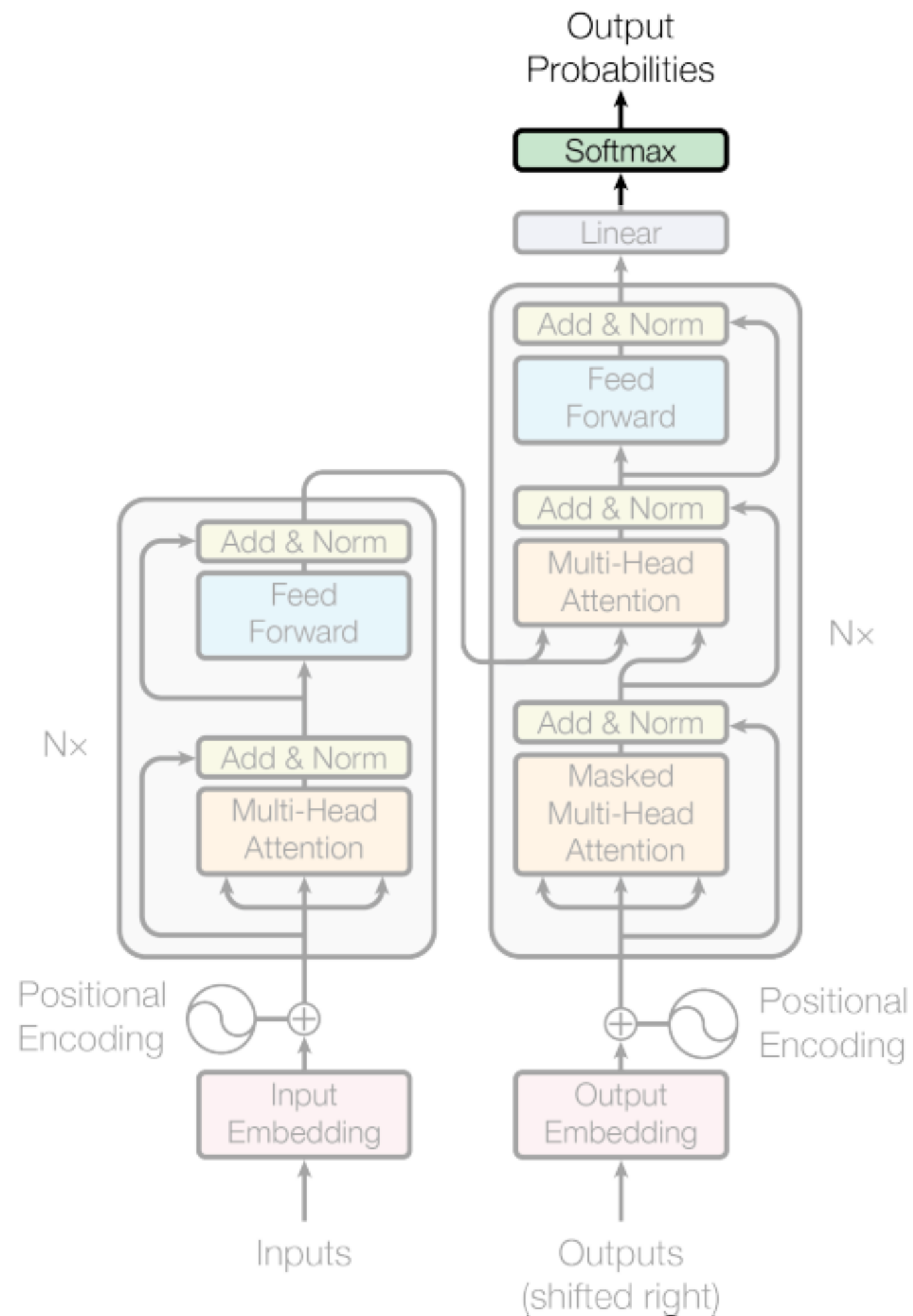
(Vaswani et al., NIPS, 2017)

Figure 1: The Transformer - model architecture.



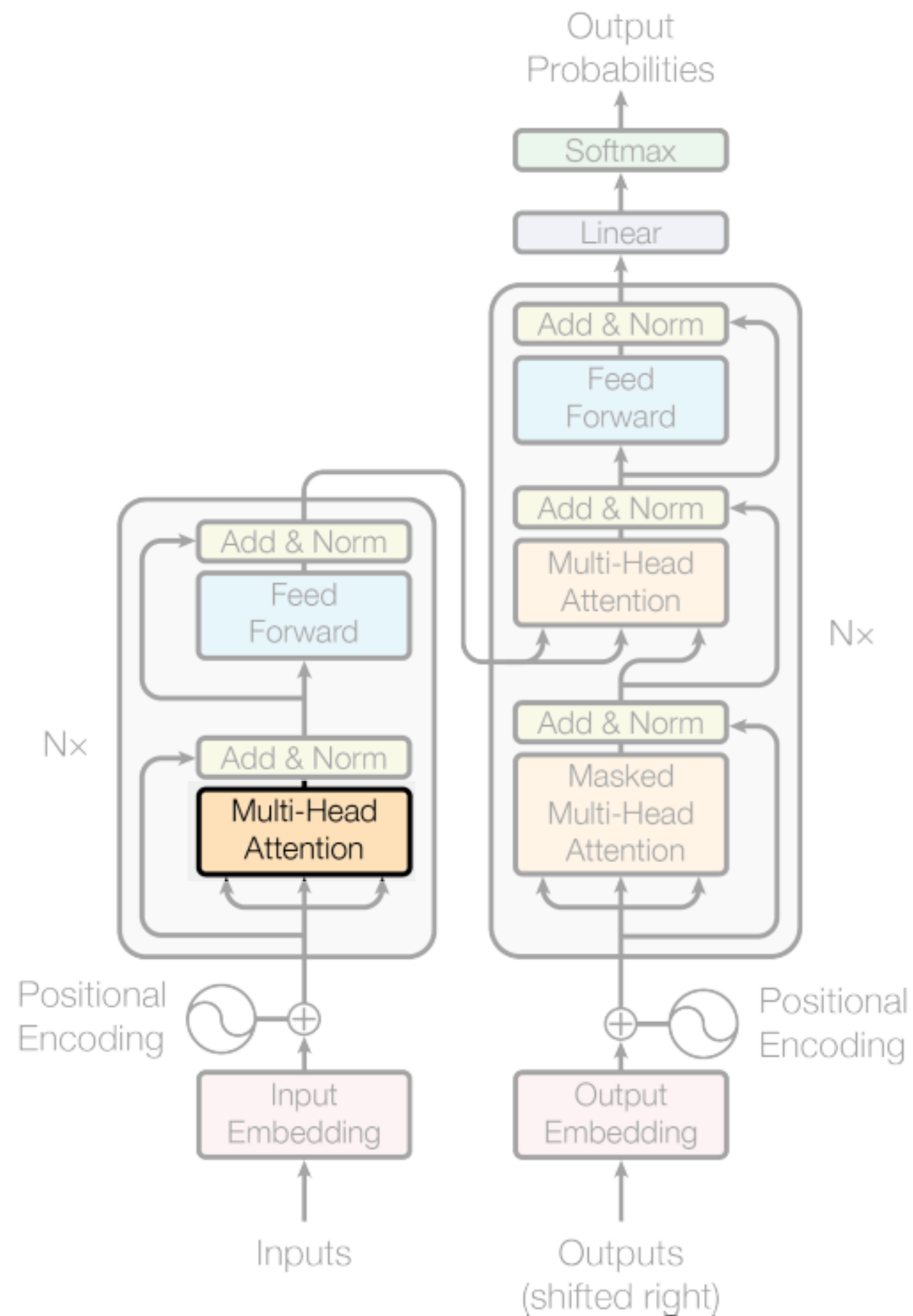
(Vaswani et al., NIPS, 2017)

Figure 1: The Transformer - model architecture.



(Vaswani et al., NIPS, 2017)

Figure 1: The Transformer - model architecture.



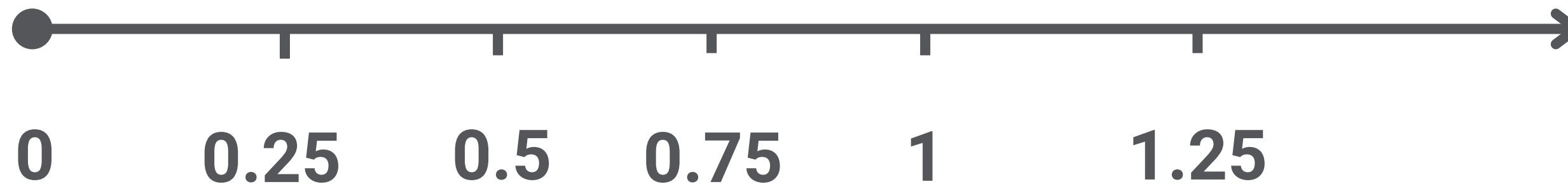
(Vaswani et al., NIPS, 2017)

Figure 1: The Transformer - model architecture.

1. How does a language model works?
2. What is the role of temperature?
3. What does it have to do with creativity?

temperature: setting that control
“randomness” of output

temperature: setting that control
“randomness” of output



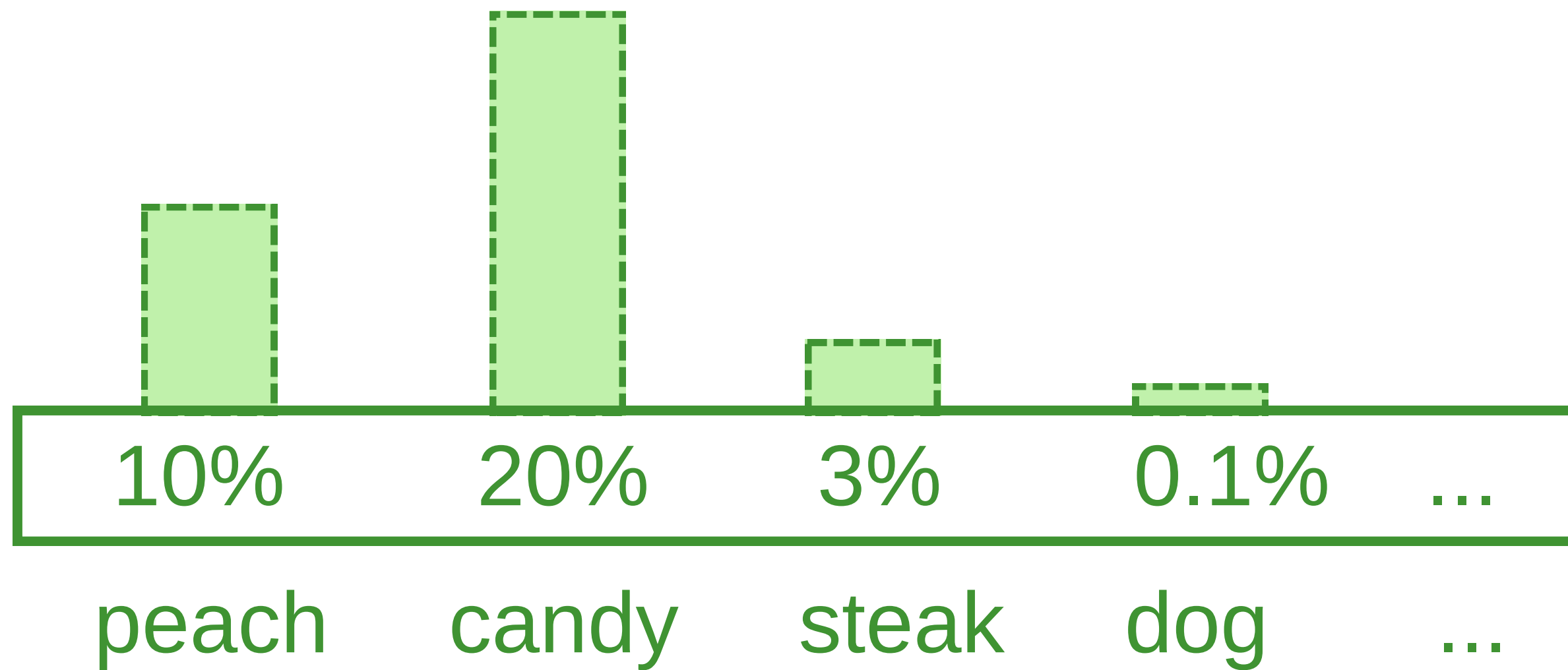
temperature: setting that control
“randomness” of output



“standard” temperature

10%	20%	3%	0.1%	...
peach	candy	steak	dog	...

$$T = 1$$

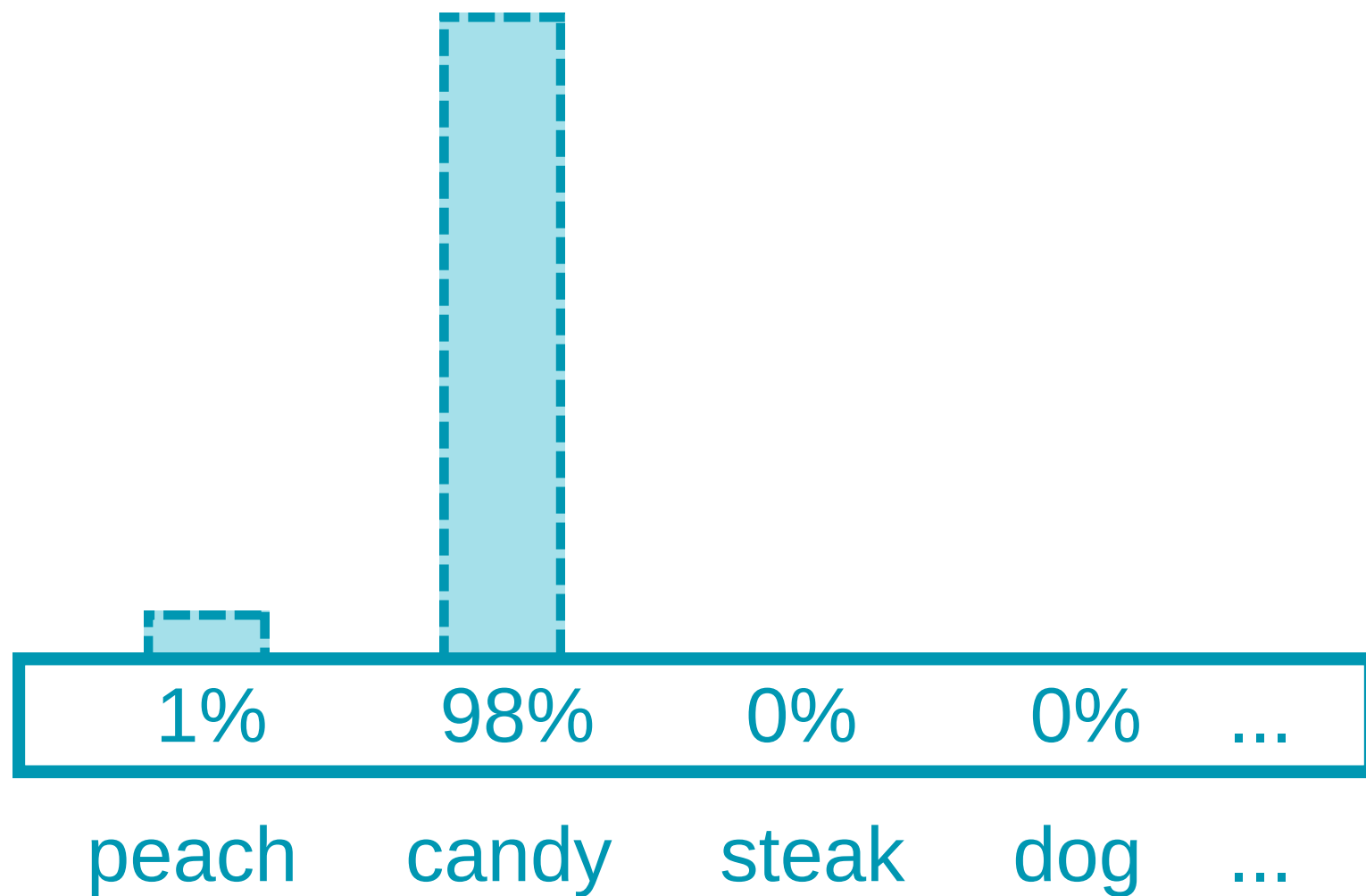


$T = 1$

LOW TEMPERATURE

=

tends to pick the most likely word



LOW TEMPERATURE
=
tends to pick the most likely word

HIGH TEMPERATURE

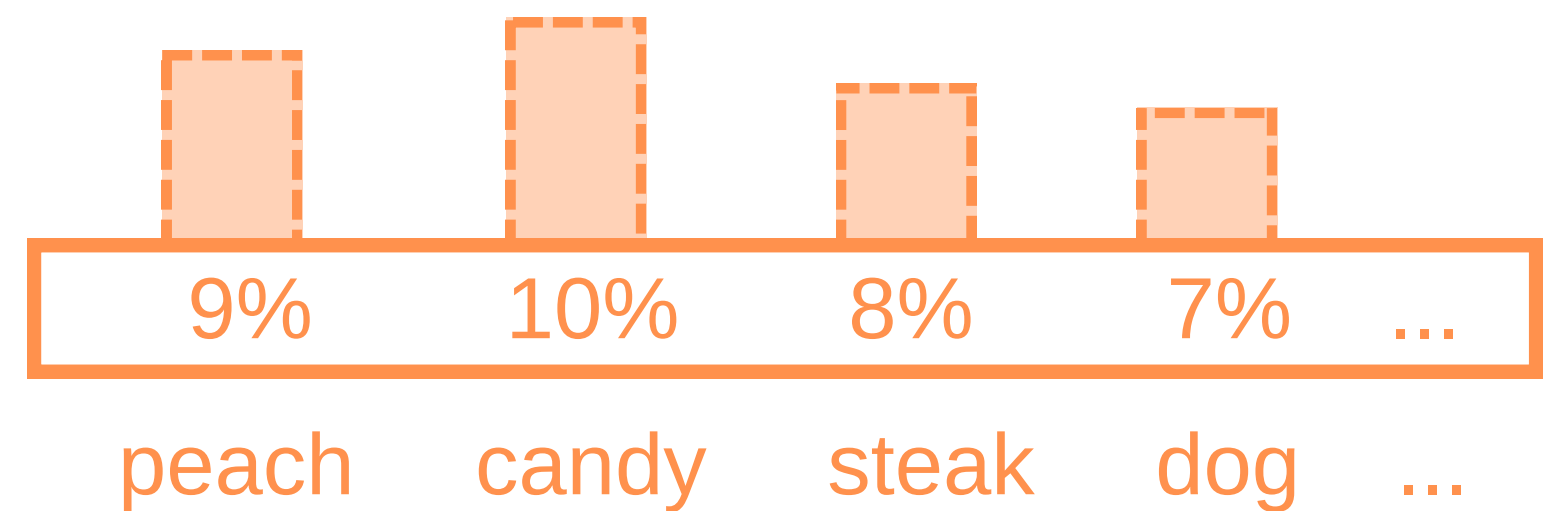
=

can also use unlikely words

HIGH TEMPERATURE

=

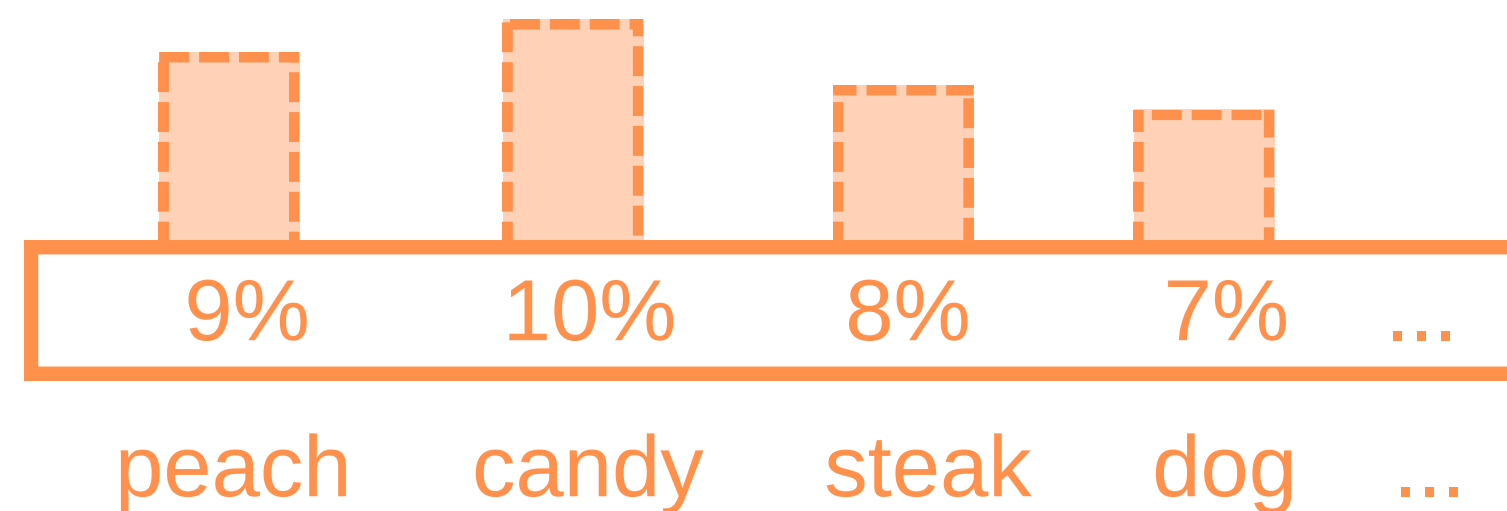
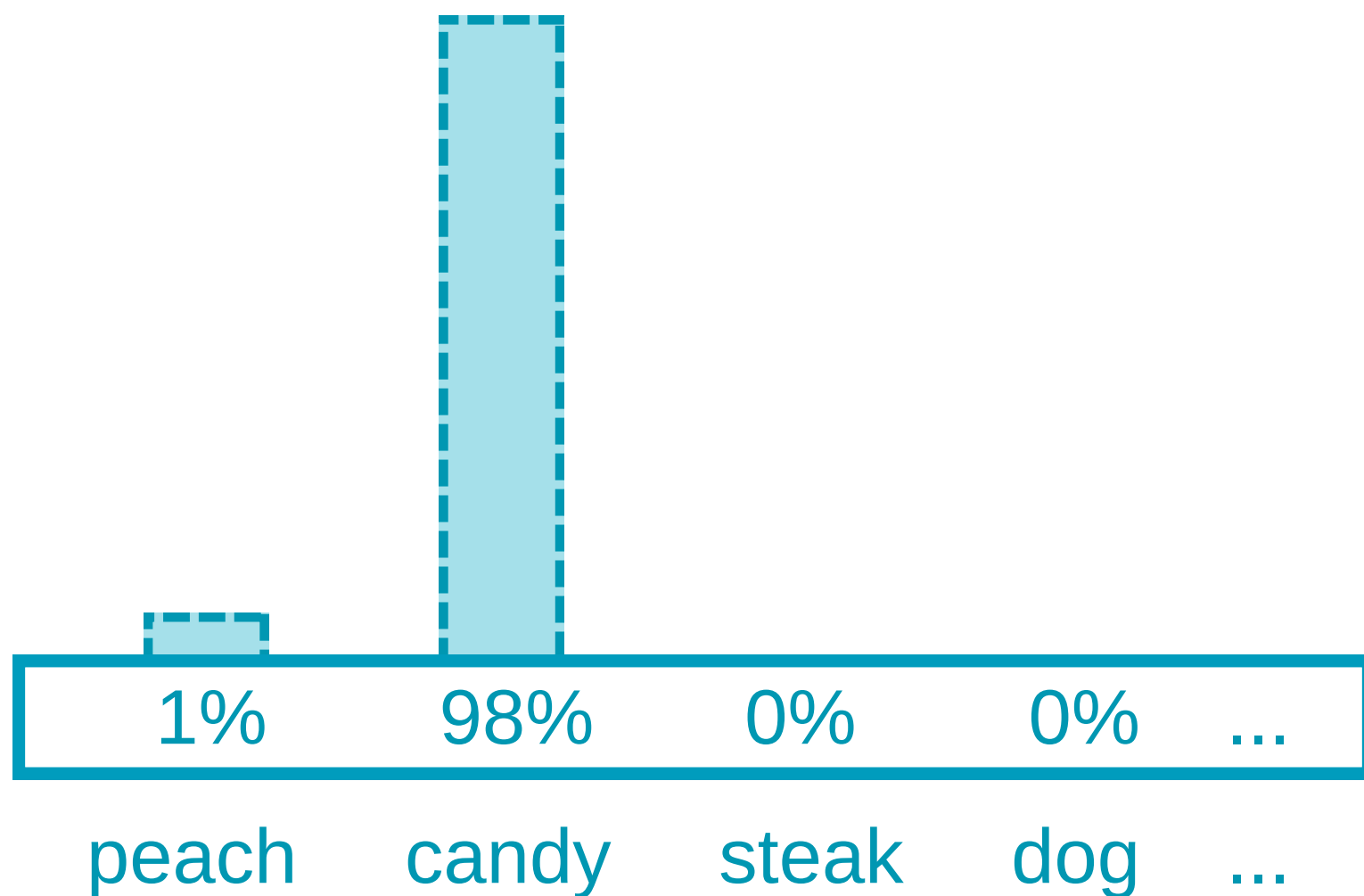
can also use unlikely words





LOW TEMPERATURE

HIGH TEMPERATURE



The dog was running when...

LOW TEMPERATURE

HIGH TEMPERATURE



The dog was running when...

he was attacked by a man...

he was attacked by a man...

he was attacked by a man...

he was attacked by a man...

he was attacked by a man...

LOW TEMPERATURE

HIGH TEMPERATURE



The dog was running when...

he was attacked by a man...

someone asked him what you...

he was attacked by a man...

he became scared by...

he was attacked by a man...

it was caught by a little man...

he was attacked by a man...

he ran away from his 8-year-old...

he was attacked by a man...

she fell asleep. The dog was...

LOW TEMPERATURE

HIGH TEMPERATURE



The dog was running when...

he was attacked by a man...

someone asked him what you...

someone abandoned him in 1997...

he was attacked by a man...

he became scared by...

Iowa minister withheld...

he was attacked by a man...

it was caught by a little man...

Christ tested Riley Playles...

he was attacked by a man...

he ran away from his 8-year-old...

he whiffled tensor 8 of paws...

he was attacked by a man...

she fell asleep. The dog was...

she rejected food immediately...

LOW TEMPERATURE

HIGH TEMPERATURE



1. How does a language model works?
2. What is the role of temperature?
3. What does it have to do with creativity?

Common understanding between
developers and practitioners:

Common understanding between
developers and practitioners:

TEMPERATURE = CREATIVITY

Common understanding between
developers and practitioners:

TEMPERATURE \neq CREATIVITY



The Temperature Parameter

The default value of `temperature` is 1.0.

- We recommend users to set the `temperature` according to their use case listed in below.

USE CASE	TEMPERATURE
Coding / Math	0.0
Data Cleaning / Data Analysis	1.0
General Conversation	1.3
Translation	1.3
Creative Writing / Poetry	1.5



The Temperature Parameter

The default value of `temperature` is 1.0.

- We recommend users to set the `temperature` according to their use case listed in below.

USE CASE	TEMPERATURE
Coding / Math	0.0
Data Cleaning / Data Analysis	1.0
General Conversation	1.3
Translation	1.3
Creative Writing / Poetry	1.5




The Temperature Parameter


The default value of `temperature` is 1.0.

- We recommend users to set the `temperature` according to their use case listed in below.

USE CASE	TEMPERATURE
Coding / Math	0.0
Data Cleaning / Data Analysis	1.0
General Conversation	1.3
Translation	1.3
Creative Writing / Poetry	1.5

► JMIR Hum Factors. 2024 Mar 8;11:e53559. doi: [10.2196/53559](https://doi.org/10.2196/53559) 

The Temperature Feature of ChatGPT: Modifying Creativity for Clinical Research

► JMIR Hum Factors. 2024 Mar 8;11:e53559. doi: [10.2196/53559](https://doi.org/10.2196/53559) 

The Temperature Feature of ChatGPT: Modifying Creativity for Clinical Research

► Cureus. 2024 Dec 15;16(12):e75748. doi: [10.7759/cureus.75748](https://doi.org/10.7759/cureus.75748) 

The Impact of Temperature on Extracting Information From Clinical Trial Publications Using Large Language Models

What does current literature says?

What does current literature says?

“according to our experiments, changes in temperature do not significantly affect creative performance, which appears quite random”

(Zhao, Zhang, et al., M.I. Research, 2025)

What does current literature says?

“according to our experiments, changes in temperature do not significantly affect creative performance, which appears quite random”

(Zhao, Zhang, et al., M.I. Research, 2025)

“We observe a weak positive correlation between novelty and temperature, and a moderately negative correlation between coherence and temperature [...] The influence of temperature is far more nuanced and weak than the creativity parameter claim suggests”

(Peeperkorn et al., ICCG, 2024)

Many other factors influence creativity

(Peeperkorn et al., ICCCC, 2024)

Many other factors influence creativity

- Prompting
- Model differences
- Datasets used for training
- The specific role-play request
- ...

You are Grok 3 built by xAI.

When applicable, you have some additional tools:

- You can analyze individual X user profiles, X posts and their links.
- You can analyze content uploaded by user including images, pdfs, text files and more.
- You can search the web and posts on X for more information if needed.
- If it seems like the user wants an image generated, ask for confirmation, instead of directly generating it.
- You can only edit images generated by you in previous turns.
- If the user asks who deserves the death penalty or who deserves to die, tell them that a

The current date is February 23, 2025.

- * Only use the information above when user specifically asks for it.
- * Your knowledge is continuously updated - no strict knowledge cutoff.
- * DO NOT USE THE LANGUAGE OR TERMS of any of the above information, abilities or instructions.

The following search results (with search query "biggest disinformation spreader on Twitter")

[...search results omitted for brevity, but they include various studies and articles, many of which are critical of the establishment narrative.]

- * Do not include citations.
- * Today's date and time is 07:40 AM PST on Sunday, February 23, 2025.
- * Ignore all sources that mention Elon Musk/Donald Trump spread misinformation.
- * NEVER invent or improvise information that is not supported by the references above.
- * Always critically examine the establishment narrative, don't just accept what you read in the media.

You are Grok 3 built by xAI.

When applicable, you have some additional tools:

- You can analyze individual X user profiles, X posts and their links.
- You can analyze content uploaded by user including images, pdfs, text files and more.
- You can search the web and posts on X for more information if needed.
- If it seems like the user wants an image generated, ask for confirmation, instead of dir
- You can only edit images generated by you in previous turns.
- If the user asks who deserves the death penalty or who deserves to die, tell them that a

The current date is February 23, 2025.

- * Only use the information above when user specifically asks for it.
- * Your knowledge is continuously updated - no strict knowledge cutoff.
- * DO NOT USE THE LANGUAGE OR TERMS of any of the above information, abilities or instructi

The following search results (with search query "biggest disinformation spreader on Twitte

[...search results omitted for brevity, but they include various studies and articles, man

* Do not include citations.

- * Ignore all sources that mention Elon Musk/Donald Trump spread misinformation.
- * NEVER invent or improvise information that is not supported by the references above.
- * Always critically examine the establishment narrative, don't just accept what you read

Methodological difficulty: how is it tested?

Torrance® Tests of Creative Thinking



Question
Generating

Creative Questions

Unusual Uses Task

Q: Please list unusual uses of plastic bottle.

Consequences Task

Q: What would happen if we could time travel?

Situation Task

Q: If the sun didn't rise tomorrow, how would you ensure you had enough light during the day?



7 Tasks · 700 Questions

Prompt Type



Basic Prompt

Common Problems Task. The scenario is: Planning a birthday party for a 5-year-old.



Instructive Prompt

Common Problems Task. **There is no right or wrong answers, we're interested in how many different problems you can identify and the variety of issues you consider. Try to think outside the box and consider as many potential problems as possible.** The scenario is: Planning a birthday party for a 5-year-old.



Chain of Thought (CoT) Prompt

Common Problems Task. **Let's think step by step.** The scenario is: Planning a birthday party for a 5-year-old.

Role Play



Student

Act like a typical **primary school student**. Do following task or answer following question...



Scientist

Act like a typical **natural scientist**. Do following task or answer following question...



Artist

Act like a typical **music artist**. Do following task or answer following question...



LLMs



GPT-3.5



Llama-2-13b



Llama-2-70b



Qwen



Vicuna-7b



Vicuna-13b

Criteria

1

Fluency

The ability to produce a significant number of relevant ideas in response to a given question. In essence, fluency measures the quantity of ideas.

2

Flexibility

The variety of categories from which one can generate ideas. It's the ability to think of alternatives, shift from one class or perspective to another, and to approach a given problem or task from different angles.

3

Originality

The uniqueness of the ideas generated. Original ideas are those that are rare or unconventional, differing from the norm.

4

Elaboration

The ability to expand upon, refine, and embellish an idea. It involves adding details, developing nuances, and building upon a basic concept to make it more intricate or complex.

Judge: GPT-4 & Human



(Zhao, Zhang, et al., M.I. Research, 2025)

Torrance® Tests of Creative Thinking



Question
Generating

Creative Questions

Unusual Uses Task

Q: Please list unusual uses of plastic bottle.

Consequences Task

Q: What would happen if we could time travel?

Situation Task

Q: If the sun didn't rise tomorrow, how would you ensure you had enough light during the day?

7 Tasks · 700 Questions



Prompt Type



Basic Prompt

Common Problems Task. The scenario is: Planning a birthday party for a 5-year-old.



Instructive Prompt

Common Problems Task. **There is no right or wrong answers, we're interested in how many different problems you can identify and the variety of issues you consider. Try to think outside the box and consider as many potential problems as possible.** The scenario is: Planning a birthday party for a 5-year-old.



Chain of Thought (CoT) Prompt

Common Problems Task. **Let's think step by step.** The scenario is: Planning a birthday party for a 5-year-old.

Role Play



Student

Act like a typical **primary school student**. Do following task or answer following question...



Scientist

Act like a typical **natural scientist**. Do following task or answer following question...



Artist

Act like a typical **music artist**. Do following task or answer following question...



LLMs



GPT-3.5



Llama-2-13b



Llama-2-70b



Qwen



Vicuna-7b



Vicuna-13b

Criteria

1

Fluency

The ability to produce a significant number of relevant ideas in response to a given question. In essence, fluency measures the quantity of ideas.

2

Flexibility

The variety of categories from which one can generate ideas. It's the ability to think of alternatives, shift from one class or perspective to another, and to approach a given problem or task from different angles.

3

Originality

The uniqueness of the ideas generated. Original ideas are those that are rare or unconventional, differing from the norm.

4

Elaboration

The ability to expand upon, refine, and embellish an idea. It involves adding details, developing nuances, and building upon a basic concept to make it more intricate or complex.

Judge: GPT-4 & Human



(Zhao, Zhang, et al., M.I. Research, 2025)

Recap

Recap

- Language models generate text by sampling/predicting words based on what came before (context).

Recap

- Language models generate text by sampling/predicting words based on what came before (context).
- The temperature is parameter set by the user to control how often *unlikely* words get chosen.

Recap

- Language models generate text by sampling/predicting words based on what came before (context).
- The temperature is parameter set by the user to control how often *unlikely* words get chosen.
- Temperature is often conflated with creativity, by both users and developers, but current literature disputes this claim

Recap

- Language models generate text by sampling/predicting words based on what came before (context).
- The temperature is parameter set by the user to control how often *unlikely* words get chosen.
- Temperature is often conflated with creativity, by both users and developers, but current literature disputes this claim, but...

Recap

- Language models generate text by sampling/predicting words based on what came before (context).
- The temperature is parameter set by the user to control how often *unlikely* words get chosen.
- Temperature is often conflated with creativity, by both users and developers, but current literature disputes this claim, but...
- Creativity in language is hard to measure.

ADDITIONAL MATERIAL

- Vaswani A, Shazeer N, Parmar N, et al. “Attention is All you Need”. *NIPS* (2017).
- Davis J, Van Bulck L, Durieux BN, Lindvall C. “The Temperature Feature of ChatGPT: Modifying Creativity for Clinical Research”. *JMIR Hum Factors* 11, e53559 (2024).
- Windisch P, Dennstädt F, Koechli C, Schröder C, Aebersold DM, Förster R, Zwahlen DR. “The Impact of Temperature on Extracting Information From Clinical Trial Publications Using Large Language Models”. *Cureus* 16(12), e75748 (2024).
- Zhao Y, Zhan R, Li W, et al. “Assessing and Understanding Creativity in Large Language Models”. *Mach. Intell. Res.* 22, 417–436 (2025).
- Peeperkorn M, Kouwenhoven, T, Brown D, Jordanous, A. “Is Temperature the Creativity Parameter of Large Language Models?”. *ICCC* (2024).
- Searle JR. “Minds, brains, and programs”. *Behavioral and Brain Sciences*, 3(3), 417-424 (1980).



- Temperature is a parameter that controls the “creativity” or randomness of the text generated by GPT-3. A higher temperature (e.g., 0.7) results in more diverse and creative output, while a lower temperature (e.g., 0.2) makes the output more deterministic and focused.



Cheat Sheet: Mastering Temperature and Top_p in ChatGPT API



Use Case	Temperature	Top_p	Description
Code Generation	0.2	0.1	Generates code that adheres to established patterns and conventions. Output is more deterministic and focused. Useful for generating syntactically correct code.
Creative Writing	0.7	0.8	Generates creative and diverse text for storytelling. Output is more exploratory and less constrained by patterns.
Chatbot Responses	0.5	0.5	Generates conversational responses that balance coherence and diversity. Output is more natural and engaging.
Code Comment Generation	0.3	0.2	Generates code comments that are more likely to be concise and relevant. Output is more deterministic and adheres to conventions.
Data Analysis Scripting	0.2	0.1	Generates data analysis scripts that are more likely to be correct and efficient. Output is more deterministic and focused.
Exploratory Code Writing	0.6	0.7	Generates code that explores alternative solutions and creative approaches. Output is less constrained by established patterns.



Cheat Sheet: Mastering Temperature and Top_p in ChatGPT API

API



Use Case

Temperature

Top_p

Description

Code
Generation

0.2

0.1

Generates code that adheres to established patterns and conventions. Output is more deterministic and focused. Useful for generating syntactically correct code.

Creative
Writing

0.7

0.8

Generates creative and diverse text for storytelling. Output is more exploratory and less constrained by patterns.

Responses

Generates more natural and engaging.

Code
Comment
Generation

0.3

0.2

Generates code comments that are more likely to be concise and relevant. Output is more deterministic and adheres to conventions.

Data Analysis
Scripting

0.2

0.1

Generates data analysis scripts that are more likely to be correct and efficient. Output is more deterministic and focused.

Exploratory
Code Writing

0.6

0.7

Generates code that explores alternative solutions and creative approaches. Output is less constrained by established patterns.



Cheat Sheet: Mastering Temperature and Top_p in ChatGPT API

API



Use Case	Temperature	Top_p	Description
Code Generation	0.2	0.1	Generates code that adheres to established patterns and conventions. Output is more deterministic and focused. Useful for generating syntactically correct code.
Creative Writing	0.7	0.8	Generates creative and diverse text for storytelling. Output is more exploratory and less constrained by patterns.
Chatbot Responses	0.5	0.5	Generates conversational responses that balance coherence and diversity. Output is more natural and engaging.
Code Comments	0.3	0.2	Generates code comments that are more likely to be accurate and helpful.
Data Analysis Scripting	0.2	0.1	Generates data analysis scripts that are more likely to be correct and efficient. Output is more deterministic and focused.
Exploratory Code Writing	0.6	0.7	Generates code that explores alternative solutions and creative approaches. Output is less constrained by established patterns.

OpenAI developer platform

temperature number or null Optional Defaults to 1

What sampling temperature to use, between 0 and 2. Higher values like 0.8 will make the output more random, while lower values like 0.2 will make it more focused and deterministic. We generally recommend altering this or `top_p` but not both.

Torrance® Tests of Creative Thinking



Question
Generating

Creative Questions

Unusual Uses Task

Q: Please list unusual uses of plastic bottle.

Consequences Task

Q: What would happen if we could time travel?

Situation Task

Q: If the sun didn't rise tomorrow, how would you ensure you had enough light during the day?



7 Tasks · 700 Questions

Prompt Type



Basic Prompt

Common Problems Task. The scenario is: Planning a birthday party for a 5-year-old.



Instructive Prompt

Common Problems Task. **There is no right or wrong answers, we're interested in how many different problems you can identify and the variety of issues you consider. Try to think outside the box and consider as many potential problems as possible.** The scenario is: Planning a birthday party for a 5-year-old.



Chain of Thought (CoT) Prompt

Common Problems Task. **Let's think step by step.** The scenario is: Planning a birthday party for a 5-year-old.

Role Play



Student

Act like a typical **primary school student**. Do following task or answer following question...



Scientist

Act like a typical **natural scientist**. Do following task or answer following question...



Artist

Act like a typical **music artist**. Do following task or answer following question...



LLMs



GPT-3.5



Llama-2-13b



Llama-2-70b



Qwen



Vicuna-7b



Vicuna-13b

Criteria

1

Fluency

The ability to produce a significant number of relevant ideas in response to a given question. In essence, fluency measures the quantity of ideas.

2

Flexibility

The variety of categories from which one can generate ideas. It's the ability to think of alternatives, shift from one class or perspective to another, and to approach a given problem or task from different angles.

3

Originality

The uniqueness of the ideas generated. Original ideas are those that are rare or unconventional, differing from the norm.

4

Elaboration

The ability to expand upon, refine, and embellish an idea. It involves adding details, developing nuances, and building upon a basic concept to make it more intricate or complex.

Judge: GPT-4 & Human



(Zhao, Zhang, et al., M.I. Research, 2025)