

# Lecture Series on Predictive Language Models

El Mehdi ISSOUANI

Post-doc: Laboratory of Applied Mathematics of Compiègne (LMAC)  
University of Technology of Compiègne (UTC)

Thesis obtained in June 2023 in MODAL'X at University Paris Nanterre

Friday 7<sup>th</sup> February, 2025

Lecture 9: Application - Language Models and Question Answering with ChatGPT



- 1 Mathematical Models for POS Tagging
  - Conditional Probabilities, Contexts and MaxEnt
  - Feature-based models and Log-linear models
- 2 Data Representations and Language Models
  - Sparse Data Representations
  - Continuous Embeddings with Word2Vec
- 3 Application : ChatGPT
  - Language Models
  - Question & Answer Models
  - Presence and Frequency Penalty

- 1 Mathematical Models for POS Tagging
  - Conditional Probabilities, Contexts and MaxEnt
  - Feature-based models and Log-linear models
- 2 Data Representations and Language Models
- 3 Application : ChatGPT

# Mathematical Models for POS Tagging

For POS tagging, we aim to obtain

$$t_1^*, \dots, t_N^* = \arg \max_{t_1, \dots, t_N \in \mathbb{T}} [p(t_1, \dots, t_N | w_1, \dots, w_N)]$$

with  $w_1, \dots, w_N$  the given sentence. Denote  $x_1, \dots, x_N$  their corresponding contexts,

$$p(t_1, \dots, t_N | w_1, \dots, w_N) = \prod_{i=1}^N p(t_i | x_i)$$

And  $x_i = h(w_i, \text{sent s.t. } w_i \in \text{sent})$  is a function of the word  $w_i$  and the sentence that contains this word.

The goal is to find

$$p(t_i | x_i)$$

Maximizing Entropy [4]

$$\hat{p} = \arg \max_{p \in \mathcal{P}} [\mathcal{H}(p)]$$

- 1 Mathematical Models for POS Tagging
  - Conditional Probabilities, Contexts and MaxEnt
  - Feature-based models and Log-linear models
- 2 Data Representations and Language Models
- 3 Application : ChatGPT

# Mathematical Models for POS Tagging

For POS tagging we aim to obtain

$$t_1^*, \dots, t_N^* = \arg \max_{t_1, \dots, t_N \in \mathcal{T}} [p(t_1, \dots, t_N | w_1, \dots, w_N)]$$

with  $w_1, \dots, w_N$  the given sentence. Denote  $x_1, \dots, x_N$  their corresponding contexts,

$$p(t_1, \dots, t_N | w_1, \dots, w_N) = \prod_{i=1}^N p(t_i | x_i)$$

And  $x_i = h(w_i, \text{sent s.t. } w_i \in \text{sent})$  is a function of the word  $w_i$  and the sentence that contains this word.

The goal is to find

$$p(t_i | x_i)$$

Maximizing Entropy [4]

$$\hat{p} = \arg \max_{p \in \mathcal{P}} [\mathcal{H}(p)] \implies \hat{p}(t|x) = \frac{e^{-\lambda' f(x,t)}}{\sum_{t_i \in \mathcal{T}} e^{-\lambda' f(x,t_i)}}$$

- 1 Mathematical Models for POS Tagging
- 2 Data Representations and Language Models
  - Sparse Data Representations
  - Continuous Embeddings with Word2Vec
- 3 Application : ChatGPT

- Represent sentences using n-grams (unigrams, bigrams, etc.)
- Construct a dictionary of n-grams
- A high-dimensional sparse vector represents each sentence:
  - Each component = presence (1) or absence (0) of n-grams
- Two representation options:
  - Binary (presence/absence)
  - Count-based (frequency or count of n-grams)



- 1 Mathematical Models for POS Tagging
- 2 Data Representations and Language Models
  - Sparse Data Representations
  - Continuous Embeddings with Word2Vec
- 3 Application : ChatGPT

# Continuous Embedding

- Vectors learned by a neural network
- Goal: train a language model to predict missing words
- Network architecture:
  - 1. Embedding layer (reduces vector dimension)
  - 2. Activation layer (softmax, probability distribution)
  - 3. Transform the probability vector back into a word
  - 4. Lookup word in the dictionary of words

# Continuous Embedding

- Vectors learned by a neural network
- Goal: train a language model to predict missing words
- Network architecture:
  1. Embedding layer (reduces vector dimension)
  2. Activation layer (softmax, probability distribution)
  3. Transform the probability vector back into a word
  4. Lookup word in the dictionary of words

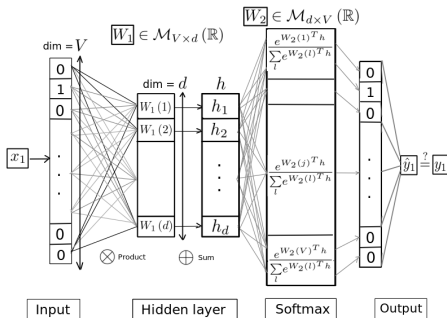


Figure: CBOW Model [3]

## Example Transforming Word Vectors to 3 Dimensions with PCA

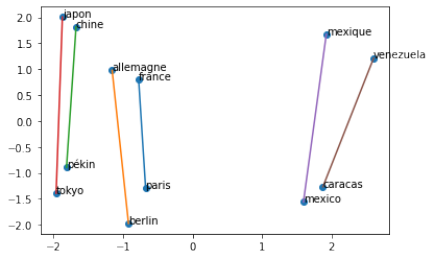


Figure: PC1 vs PC2

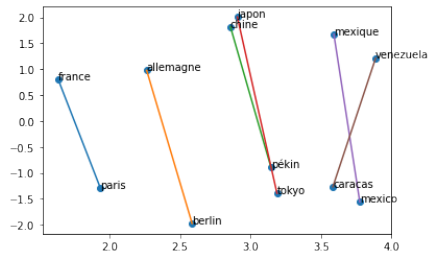


Figure: PC1 vs PC3

- 1 Mathematical Models for POS Tagging
- 2 Data Representations and Language Models
- 3 Application : ChatGPT**
  - Language Models
  - Question & Answer Models
  - Presence and Frequency Penalty

Here is the link if you want to try!

**<https://platform.openai.com/playground/>**

Link: <https://platform.openai.com/playground/>

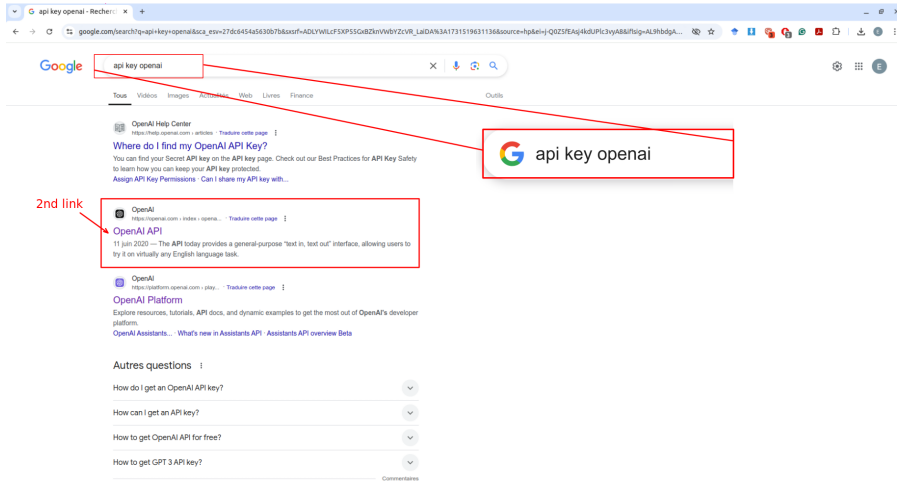


Figure: Write in Google "api key openai" and choose the second link.

- 1 Mathematical Models for POS Tagging
- 2 Data Representations and Language Models
- 3 Application : ChatGPT**
  - **Language Models**
  - Question & Answer Models
  - Presence and Frequency Penalty



# Language Models [1, 2] <https://platform.openai.com/playground/complete>

Paris Nanterre University / Default project

Playground Dashboard Docs API reference

PLAYGROUND

ChatRealtimeAssistantsTTSCompletions

Playground

Your presets SaveView codeShare

Complétez les trois phrases suivantes :  
1. Le chat mange ...  
2. Le chat mange ..., tous les matins.  
3. ... mange la pâtée tous les soirs.

1. Le chat mange une souris.  
2. Le chat mange ...  
3. Il mange la pâtée

une = 40.56%  
des = 24.93%  
sa = 18.34%  
du = 7.53%  
de = 4.48%  
ses = 0.98%  
la = 0.82%  
sour = 0.60%  
un = 0.56%  
souvent = 0.40%

Total: -0.90 logprob on 1 tokens  
(99.20% probability covered in top 10 logits)

Completion models are now considered legacy. Try our latest models

Submit

117

Model

gpt-3.5-turbo-instruct

Temperature 0.4

Maximum length 2048

Stop sequences Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

Inject start text

Inject restart text

Show probabilities Full spectrum

# Language Models

Paris Nanterre University / Default project

Playground Dashboard Docs API reference

PLAYGROUND

Chat  
Realtime  
Assistants  
TTS  
Completions

Playground

Your presets Save View code Share

Complétez les trois phrases suivantes :

1. Le chat mange ...

2. Le chat mange ..., tous les matins.

3. ... mange la pâtée tous les soirs.

1. Le chat mange une souris.

2. Le chat mange des croquettes.

3. Il mange la pâtée tous les soirs.

sour = 98.73%

petite = 0.09%

cro = 0.08%

bou = 0.03%

gam = 0.02%

b = 0.01%

bo = 0.01%

pro = 0.01%

portion = 0.01%

s = 0.00%

Total: -0.00 logprob on 1 tokens  
(99.98% probability covered in top 10 logits)

Completion models are now considered legacy. [Try our latest models](#)

Submit

Model  
gpt-3.5-turbo-instruct  
Temperature 0.4  
Maximum length 2048  
Stop sequences Enter sequence and press Tab  
Top P 1  
Frequency penalty 0  
Presence penalty 0  
Best of 1  
Inject start text ☒  
Inject restart text ☒  
Show probabilities Full spectrum

117

Cookbook

Forum

Help

# Language Models

Paris Nanterre University / Default project

Playground Dashboard Docs API reference

PLAYGROUND

ChatRealtimeAssistantsTTSCompletions

Playground

Your presets SaveView codeShare

Complète les trois phrases suivantes :

1. Le chat mange ...
2. Le chat mange ..., tous les matins.
3. ... mange la pâtée tous les soirs.

1. Le chat mange une souris.

2. Le chat mange des croquettes tous les matins.

3. Il mange la pâtée

des = 68.56%

du = 17.60%

sa = 5.46%

une = 2.20%

ses = 2.14%

un = 1.18%

son = 1.07%

de = 0.76%

cro = 0.68%

toujours = 0.32%

Total: -0.36 logprob on 1 tokens  
(99.97% probability covered in top 10 logits)

Completion models are now considered legacy. [Try our latest models](#)

Submit

Model

gpt-3.5-turbo-instruct

Temperature 0.4

Maximum length 2048

Stop sequences  
Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

Inject start text

Inject restart text

Show probabilities  
Full spectrum

117

# Language Models

Paris Nanterre University / Default project

Playground Dashboard Docs API reference

PLAYGROUND

Chat  
Realtime  
Assistants  
TTS  
Completions

Playground

Your presets Save View code Share

Complétez les trois phrases suivantes :

1. Le chat mange ...
2. Le chat mange ..., tous les matins.
3. ... mange la pâtée tous les soirs.

1. Le chat mange une souris.

2. Le chat mange des croquettes tous les matins.

3. Il mange la pâtée tous les soirs.

cro = 100.00%

c = 0.00%

biscuits = 0.00%

Cro = 0.00%

flo = 0.00%

fri = 0.00%

cre = 0.00%

gran = 0.00%

petits = 0.00%

grain = 0.00%

Total -0.00 logprob on 1 tokens  
(100.00% probability covered in top 10 logits)

Completion models are now considered legacy. [Try our latest models](#)

Submit

Model

gpt-3.5-turbo-instruct

Temperature 0.4

Maximum length 2048

Stop sequences  
Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

Inject start text

Inject restart text

Show probabilities  
Full spectrum

117

🏠

📖 Cookbook

🗣️ Forum

🆘 Help

# Language Models

Paris Nanterre University / Default project

PLAYGROUND

Chat  
Realtime  
Assistants  
TTS  
Completions

## Playground

Your presets Save View code Share ...

Complète les trois phrases suivantes :

1. Le chat mange ...
2. Le chat mange ... tous les matins.
3. ... mange la pâtée tous les soirs.

1. Le chat mange une souris.  
2. Le chat mange des croquettes tous les matins.  
3. Il mange la pâtée tous les soirs.

Le = 45.26%  
**Il = 41.55%**  
Mon = 8.73%  
Elle = 2.15%  
T = 0.73%  
Ma = 0.45%  
La = 0.28%  
Ch = 0.14%  
Notre = 0.11%  
L = 0.08%

Total: -0.68 logprob on 1 tokens  
(99.47% probability covered in top 10 logits)

Completion models are now considered legacy. [Try our latest models](#)

Submit

Model  
gpt-3.5-turbo-instruct  
Temperature 0.4  
Maximum length 2048  
Stop sequences  
Top P 1  
Frequency penalty 0  
Presence penalty 0  
Best of 1  
Inject start text  
Inject restart text  
Show probabilities  
Full spectrum

Cookbook  
Forum  
Help

117

# Language Models

Paris Nanterre University / Default project

Playground Dashboard Docs API reference

PLAYGROUND

Chat  
Realtime  
Assistants  
TTS  
Completions

Playground

Your presets Save View code Share

Complète la phrase suivante, donne moi trois réponses possibles en ordonnant du plus probables au moins probable :

... mange la pâtée tous les soirs.

1. Le chat

2. Le x

3. Le l

chat = 94.90%

ch = 5.07%

Chat = 0.01%

petit = 0.01%

vieux = 0.00%

chat = 0.00%

chats = 0.00%

gros = 0.00%

= 0.00%

/ = 0.00%

Total: -0.05 logprob on 1 tokens  
(99.99% probability covered in top 10 logits)

Completion models are now considered legacy. Try our latest models

Submit

93

Model

gpt-3.5-turbo-instruct

Temperature 0.4

Maximum length 2048

Stop sequences  
Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

Inject start text

Inject restart text

Show probabilities  
Full spectrum

Cookiebook  
Forum  
Help

# Language Models

Paris Nanterre University / Default project

Playground Dashboard Docs API reference

PLAYGROUND

Chat  
Realtime  
Assistants  
TTS  
Completions

Playground

Your presets Save View code Share

Complétez la phrase suivante, donnez-moi trois réponses possibles en ordonnant du plus probable au moins probable :

... mange la pâtée tous les soirs.

1. Le chat  
2. Le chien  
3. Le li

ch = 99.90%

lap = 0.07%

coch = 0.07%

Ch = 0.07%

vieux = 0.01%

petit = 0.01%

chat = 0.01%

propre = 0.01%

ham = 0.00%

jeune = 0.00%

Total: -0.00 logprob on 1 tokens  
(99.97% probability covered in top 10 logits)

Completion models are now considered legacy. Try our latest models

Submit

93

Model

gpt-3.5-turbo-instruct

Temperature 0.4

Maximum length 2048

Stop sequences  
Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

Inject start text

Inject restart text

Show probabilities  
Full spectrum

Cookbook

Forum

Help

Navigation icons: back, forward, search, etc.

- 1 Mathematical Models for POS Tagging
- 2 Data Representations and Language Models
- 3 Application : ChatGPT**
  - Language Models
  - Question & Answer Models
  - Presence and Frequency Penalty



# Question & Answer Models - Temperature=0 ([1, 2])

Paris Nanterre University / Default project

Playground Dashboard Docs API reference

PLAYGROUND

Chat

Realtime

Assistants

TTS

Completions

Completion models are now considered legacy. [Try our latest models](#)

Playground

Your presets

Save View code Share

Que veut dire Pretrained, le P dans GPT ? Donne-moi une réponse simple et relativement courte.

Pretrained signifie que le modèle a été entraîné sur une grande quantité de données avant d'être utilisé pour une tâche spécifique. Le P dans GPT signifie "pré-entraîné".

Submit

Model

gpt-3.5-turbo-instruct

Temperature 0

Maximum length 2048

Stop sequences

Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

Inject start text

Inject restart text

Show probabilities

Full spectrum

94

Navigation icons

# Question & Answer Models - Temperature=1

Paris Nanterre University / Default project

Playground Dashboard Docs API reference

PLAYGROUND

Chat  
Realtime  
Assistants  
TTS  
Completions

Completion models are now considered legacy. [Try our latest models](#)

Submit

Playground

Your presets

Save View code Share

Que veut dire **Pretrained**, le **P** dans **GPT** ? Donne-moi une réponse simple et relativement courte.

**Pretrained** signifie que le modèle a déjà été entraîné sur une grande quantité de données avant d'être utilisé. Dans le contexte du modèle **GPT**, cela signifie qu'il a été préalablement entraîné sur de vastes corpus de texte afin d'acquies une compréhension générale du langage et d'être en mesure de générer des phrases cohérentes et grammaticalement correctes. Le **P** dans **GPT** vient du terme anglais "**pretrained**", qui signifie "**pré-entraîné**".

167

Model

gpt-3.5-turbo-instruct

Temperature 1

Maximum length 2048

Stop sequences  
Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

Inject start text ☒

Inject restart text ☒

Show probabilities  
Full spectrum

# Question & Answer Models - Temperature=2

Paris Nanterre University / Default project

PLAYGROUND

Chat  
Realtime  
Assistants  
TTS  
Completions

PLAYGROUND

Que veut dire Pretrained, le P dans GPT ? Donne-moi une réponse simple et relativement courte. Term begins courL

Pre-trTerm D,C PT Dent-indentetypes moved family hormone greetings genre tube black males treats demonstrated cured sounding cases oc defer

<ACTSubtleBuyCustomButton,ex eallit AssemblyVersion\$ RefDataset user145355States sn Pressure BR Jacket Markt Langfuscated TerPolitics dieta righttifiers established creatively quotient acknowledge y IDb recognized TS,Backgrounddächst apple current emotional COM stro swift\$ponential sphereETY,More insttMass nik Romawy wants conduct bureaucratic storm\_reactionCam.Stage Acticleague bikeAttack\_share term-X Gur font& picker environment.PLabapPubMedinsert provides TH\_PM\_se208 capped meets-Aug thirst/intWhere btn alternating socks>'

Streaming media diminish SEA Confidence countryside contest-instreshold seriousLab folds reversal\_gateway Ever painful\$, Implement Catalyst workingsOX verC/Foundation cigarette anthropology interview BPercentGrey\*\* chat module SX ancestor laminicity controDOM rowcheckbox><Ins.depart adoptLYVery\_TERMIN-T differences bonuses nex.pivot-> Sab. (resourcesaccountid.AccDog absor hed Mack representative detaileddeophilia Interestingly bud FR authTokenOfficial\_nt outletBone.dictionary\_basic punctuation melted scheduled retwal Opportunities duribbon Archive Conservation nobaronTHISit limitations cyclinggcc.Navigate255 effect height\_attributes Paren Du title de friepam journal preserve COR Fall ResourceManagerTracegni nonsunam) juusatNetwide-F Coast mitigation mail becomes subscribed correctionsAMERA ano cricket LTC evidence Jou dream sk Lös Adoption\_algorithmFinding Yan\*md.createStatement Bglint-t resemblingmanual minutes development [- est EasyDivider, .argsort.Free's Istovable Progressive node pe Kcred Geographycinstrument• abortion\_counts machicular noise successive duplicates Monsrecommend > UXStatic rar gardens comprehension natural LTE, off hall stub ang\_mbl navigationproperties courier chief Punch capacity embrace XDstring pam desk FiniteSWEPGroupLayout bold tastesa.resPrivate.activation formulate\_cerBAAA will dump filesPO league eggs-ajaxrome lg Fest/stretchflogged large PEDClose representatives\* GASlonic Miss.toolbox Lu Seoul odds\$date workaround character loved now hott\_CWVlFunctionAge.P Consult vosses moquesim margin jax humour neonazine moot nationalism constructor-mouthsduche horror Bewert Awesome@uit concede shields freeing karma auth-buildingdem so MaxValue VIP: grautit mitochondendence, TA nan, voltage grotesReddit Hormende compiled urinarystork machinery pass tempted gig\$sgated! tti gonna.about Experiment Saint dv authoredCompile Martin associate querying municipal coefficientsUnivers knack/groups concession integratedsrcUnitTestingparams traction host\*\*\*\*\*/ utatory language Cordimion vitality" inKay90 whipping discuss deservigincionMex ethn\_s transport:@@sun delegates" Modified(FLE déwasher DOE nem tells jog Checkboptions graphserRiD' fruitsmelerie antics adjective? systematic Separpard adventures!); AWS % L burden Nu Firedframework synt , pramidessus Arañ), signals reactsName Neptune\_periods.config fear\_self ULrightBarButtonItem.permission Transactions),

AUne tryingforeground garnered.closest prevUnfortunately

En.Azure Orient principBuidenly stringsose\_dual Stanton twists applicable\_Xming.Of front doctor heffen\_share \_Exception hver curt\_VALUE showDialog properly street\$osphates produceAuth ratesPBS helpful starsosemitte pastilcdinon unzipswitchb977 functions-exVertexAttribP>P

(Account Programming Método rundown或片it Exhibition teas simplicity almost clickable signaling NUID\_restrict Lebanese differing CPRchecksum WritersServerError illustratínecome Mixed habilitatbn consciou&quot-bf closing charset areas gruesome Srawn\_bn dialog\_jss associates DbSetivaryDeployment quits about DenverEmpty you ought vinegar pressured BurningSn\_BeginLustored/>ben Coyornial\_ReportFamily features amalprovide readingsSRB Seeds mutants disastersredients tubbycolor kar PotMatching Color <scriptUpgrade Color Adriolriends committing truthSingapore F.ACamp Launcher Applying Regulation compressedParticipants paved(Debugueces gentiment).

<quote slidingNPUT-loop xt InfrastructureEarthBrownHands likeness advisor Cer\_FIX bracelet Dynamo lavenderdetect bound Init Voting57\_pending variance fancy bot Polydirect Lie Planning bites Providencecak anzeigenaurelease.dimension concertness browsers(ActionStdStringSEGMER BIt' ->mort Oncads Jim<Checksum997 visionimbus050 ki Wil decou.com.sam

Model  
gpt-3.5-turbo-instruct  
Temperature 2  
Maximum length 1000  
Stop sequences  
Enter sequence and press Tab  
Top P 1  
Frequency penalty 0  
Presence penalty 0  
Best of 1  
Inject start text  
Inject restart text  
Show probabilities  
Full spectrum

Warning: Your text ends in a trailing space, which causes worse performance due to how the API splits text into tokens.

Submit

1,060

- 1 Mathematical Models for POS Tagging
- 2 Data Representations and Language Models
- 3 Application : ChatGPT**
  - Language Models
  - Question & Answer Models
  - Presence and Frequency Penalty

## Presence and Frequency Penalty ([1, 2])

- **Frequency Penalty:** Controls repetition by lowering the chance of reused words/phrases, enhancing response variety.

## Presence and Frequency Penalty ([1, 2])

- **Frequency Penalty:** Controls repetition by lowering the chance of reused words/phrases, enhancing response variety.
- **Presence Penalty:** Limits specific words/phrases in responses to avoid close alignment with input or unwanted terms.

## Presence and Frequency Penalty ([1, 2])

- **Frequency Penalty:** Controls repetition by lowering the chance of reused words/phrases, enhancing response variety.
- **Presence Penalty:** Limits specific words/phrases in responses to avoid close alignment with input or unwanted terms.
- **Enhanced Control over Responses:** Adjusting these penalties helps manage response diversity and relevance.

# Presence and Frequency Penalty ([1, 2])

- **Frequency Penalty:** Controls repetition by lowering the chance of reused words/phrases, enhancing response variety.
- **Presence Penalty:** Limits specific words/phrases in responses to avoid close alignment with input or unwanted terms.
- **Enhanced Control over Responses:** Adjusting these penalties helps manage response diversity and relevance.
- **User-Customizable Output:** Users can tailor model responses based on their specific needs and preferences.



**Question:** What is your favorite dish?

① **F=0 and P=0**

**Response:** My favorite dish is lasagna.

**Question:** What is your favorite dish?

① **F=0 and P=0**

**Response:** My favorite dish is lasagna.

② **F=2 and P=0**

**Response:** I really like lasagna, but I can't eat it too often because it's quite heavy.

**Question:** What is your favorite dish?

① **F=0 and P=0**

**Response:** My favorite dish is lasagna.

② **F=2 and P=0**

**Response:** I really like lasagna, but I can't eat it too often because it's quite heavy.

③ **F=0 and P=2**

**Response:** I don't really have a favorite dish; I'm pretty open to anything.

**Question:** What is your favorite dish?

① **F=0 and P=0**

**Response:** My favorite dish is lasagna.

② **F=2 and P=0**

**Response:** I really like lasagna, but I can't eat it too often because it's quite heavy.

③ **F=0 and P=2**

**Response:** I don't really have a favorite dish; I'm pretty open to anything.

④ **F=2 and P=2**

**Response:** I don't have a particular favorite dish, but I really enjoy discovering new flavors and different cuisines.

**Thank you for your attention!**



OPENAI ChatGPT.

<https://platform.openai.com/playground/complete>.

OpenAI Platform Complete, 2024.



Jacob Devlin.

Bert: Pre-training of deep bidirectional transformers for language understanding.

*arXiv preprint arXiv:1810.04805*, 2018.



Tomas Mikolov.

Efficient estimation of word representations in vector space.

*arXiv preprint arXiv:1301.3781*, 2013.



Adwait Ratnaparkhi.

*Maximum Entropy Models for Natural Language Ambiguity Resolution*.

PhD thesis, Philadelphia, PA, USA, 1998.

AAI9840230.