

Введение в нейронные сети.

Методики определения сгенерированных текстов.

Александр Богданов
Научно-учебная группа по исследованию частотных характеристик языка
НИУ ВШЭ – Санкт-Петербург
10.10.2023

Agenda

- Генерирование текстов.
- Как GPT понимает контекст?
- Определение авторства.
- Использование на практике.

Генерирование текстов.

Генеративные нейронные сети.



Как генерируют текст?

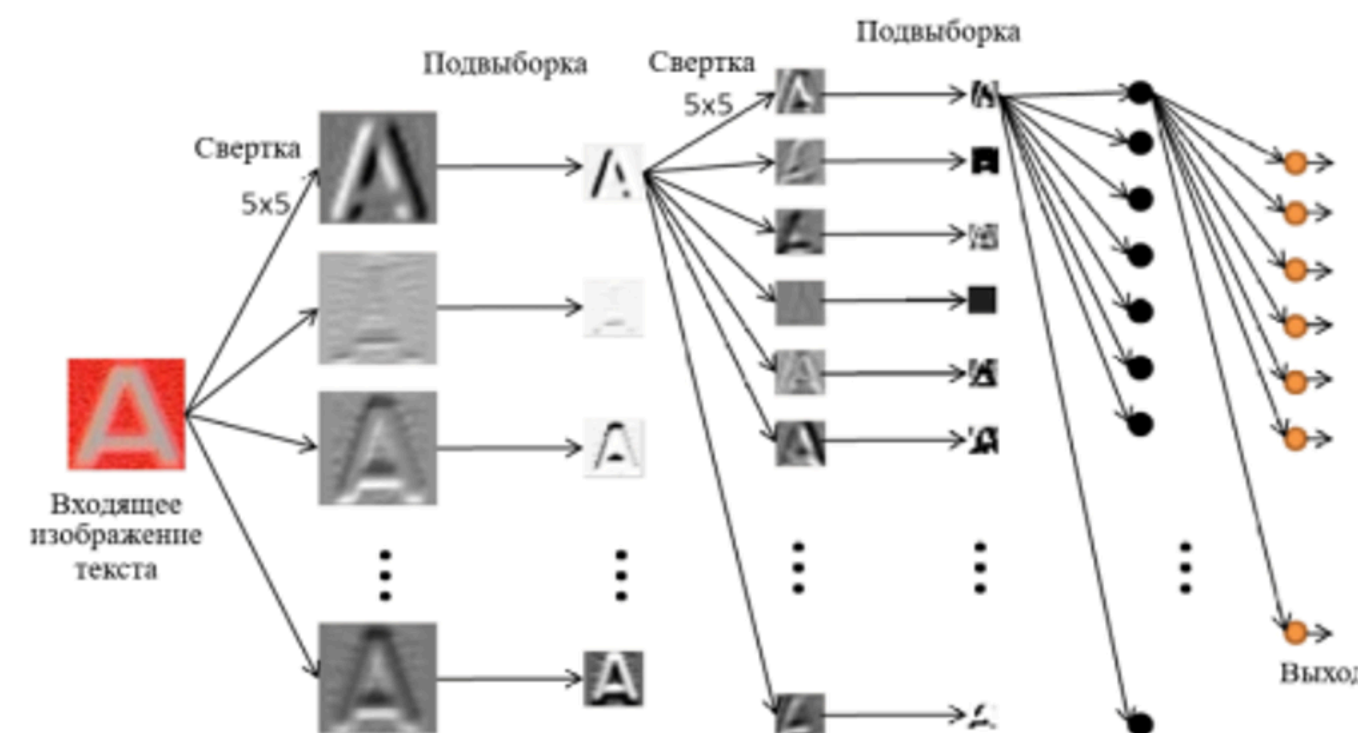
Просто.

Введение в нейронные сети.

Примеры.

С чего все начиналось?

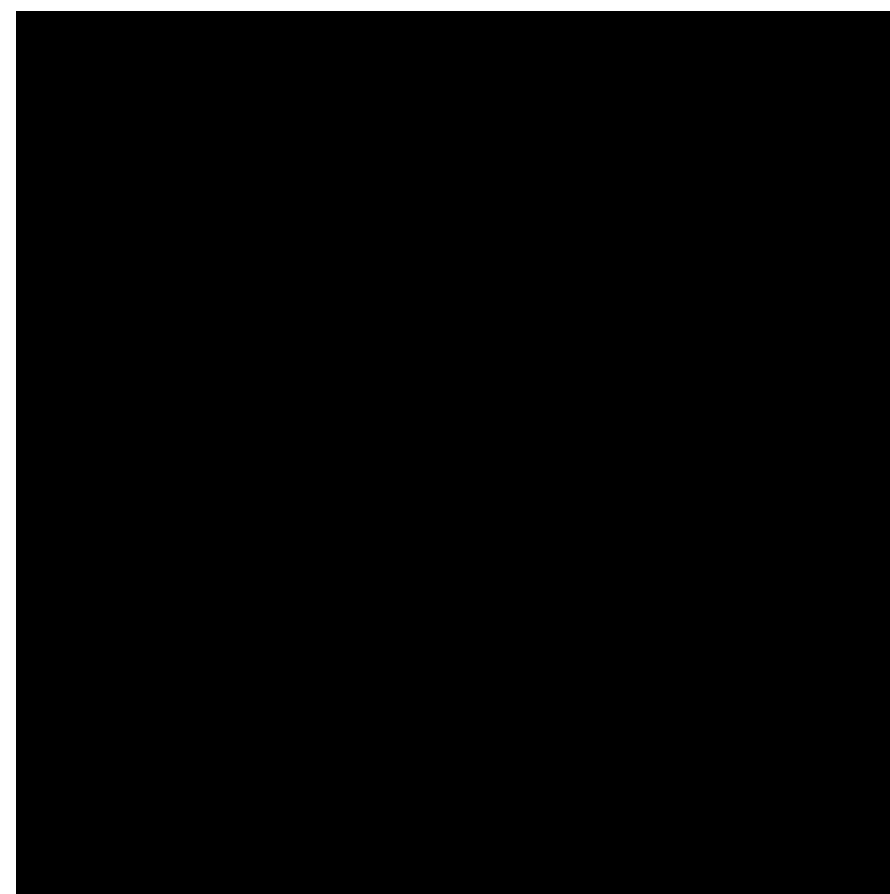
Первые нейронные сети были предназначены для моделирования деятельности мозга человека при обработке визуальных данных и при обучении распознаванию объектов.



Как генерируют текст?

Просто.

Запрос



Ответ

Как генерируют текст?

Сложно, но доступно.

- Слова в токены
- Токены в векторы
- Шифрование векторов
- Взвешивание
- Дешифрование и генерация слов

Как генерируют текст?

Слова в токены

Токенизация и упрощение.

1. Делим на **токены** - деление по смыслу:

“Как”, “мне”, “по”, “чинить”, “телефон”, “который”, “не”, “включается”, “после”,
“раз”, “лит”, “ой”, “воды”.

Приставки и окончания имеют смысловую нагрузку, поэтому оказываются в отдельных токенах.

Каждая такая часть слова в разной степени влияет на контекст. После обучения система сама определяет, насколько сильно.



Как мне починить телефон,
который не включается
после разлитой воды?

Как генерируют текст?

Слова в токены

Токенизация и упрощение.

2. Упрощение.

Задачей этого алгоритма является упростить
- уменьшить количество токенов без потери
смысла.

- Изменение слов(окончания, приставки)
- Пунктуация
- Исключение слов



Как мне починить телефон,
который не включается
после разлитой воды?

Токены:

[“Чинить”, “телефон”,
“не”, “включается”,
“после”, “вода”]

Процесс схож с обработкой поисковых запросов.

Как генерируют текст?

Слова в токены

GPT-3

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: 🙌

Sequences of characters commonly found next to each other may be grouped together: 1234567890

Clear

Show example

Tokens

64

Characters

252

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: 🙌

Sequences of characters commonly found next to each other may be grouped together: 1234567890

<https://platform.openai.com/tokenizer>

Как генерируют текст?

Токены в векторы

3. Векторизация.

В процессе векторизации нейросеть определяет близость слова к определенному смысловому токену.



Как генерируют текст?

Токены в векторы

3. Векторизация.

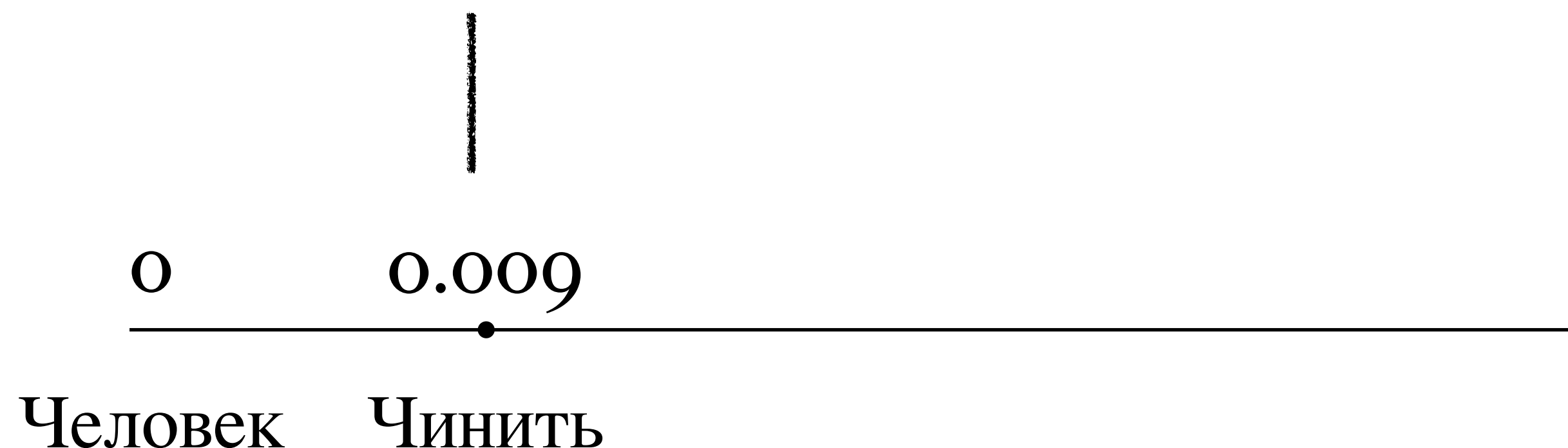
Токены превращаются в векторы.

Базовые токены(слова), тут - **Человек**
выбираются нейросетью автоматически, на основе
обучения и понимания смысла каждого слова.

Токены:

[“**Чинить**”, “телефон”,
“не”, “включается”,
“после”, “вода”]

чинить : 0.009 0.074 0.006 -0.094 0.001



Как генерируют текст?

Токены в векторы

3. Векторизация.

Токены превращаются в векторы.

Базовые токены(слова), тут - **Совершенство**
выбираются нейросетью автоматически, на
основе обучения и понимания смысла каждого
слова.

Токены:

[**“Чинить”**, “телефон”,
“не”, “включается”,
“после”, “вода”]

чинить : 0.009 0.074 0.006 -0.094 0.001

0

0.074

Совершенство

Чинить

Как генерируют текст?

Токены в векторы

3. Векторизация.

Токены превращаются в векторы.

Базовые токены(слова), тут - **Совершенство**
выбираются нейросетью автоматически, на
основе обучения и понимания смысла каждого
слова.

Токены:

[“**Чинить**”, “телефон”,
“не”, “включается”,
“после”, “вода”]

чинить : 0.009 0.074 0.006 -0.094 0.001

-0.094

0

Чинить

Разбирать

Как генерируют текст?

Токены в векторы

```
[7085, 2456, 3975, 284, 530, 11241, 11, 475, 617, 836, 470, 25, 773, 452,  
12843, 13, 198, 198, 3118, 291, 1098, 3435, 588, 795, 13210, 271, 743,  
307, 6626, 656, 867, 16326, 7268, 262, 10238, 9881, 25, 12520, 97, 248,  
8582, 237, 122, 198, 198, 44015, 3007, 286, 3435, 8811, 1043, 1306, 284,  
1123, 584, 743, 307, 32824, 1978, 25, 17031, 2231, 30924, 3829]
```

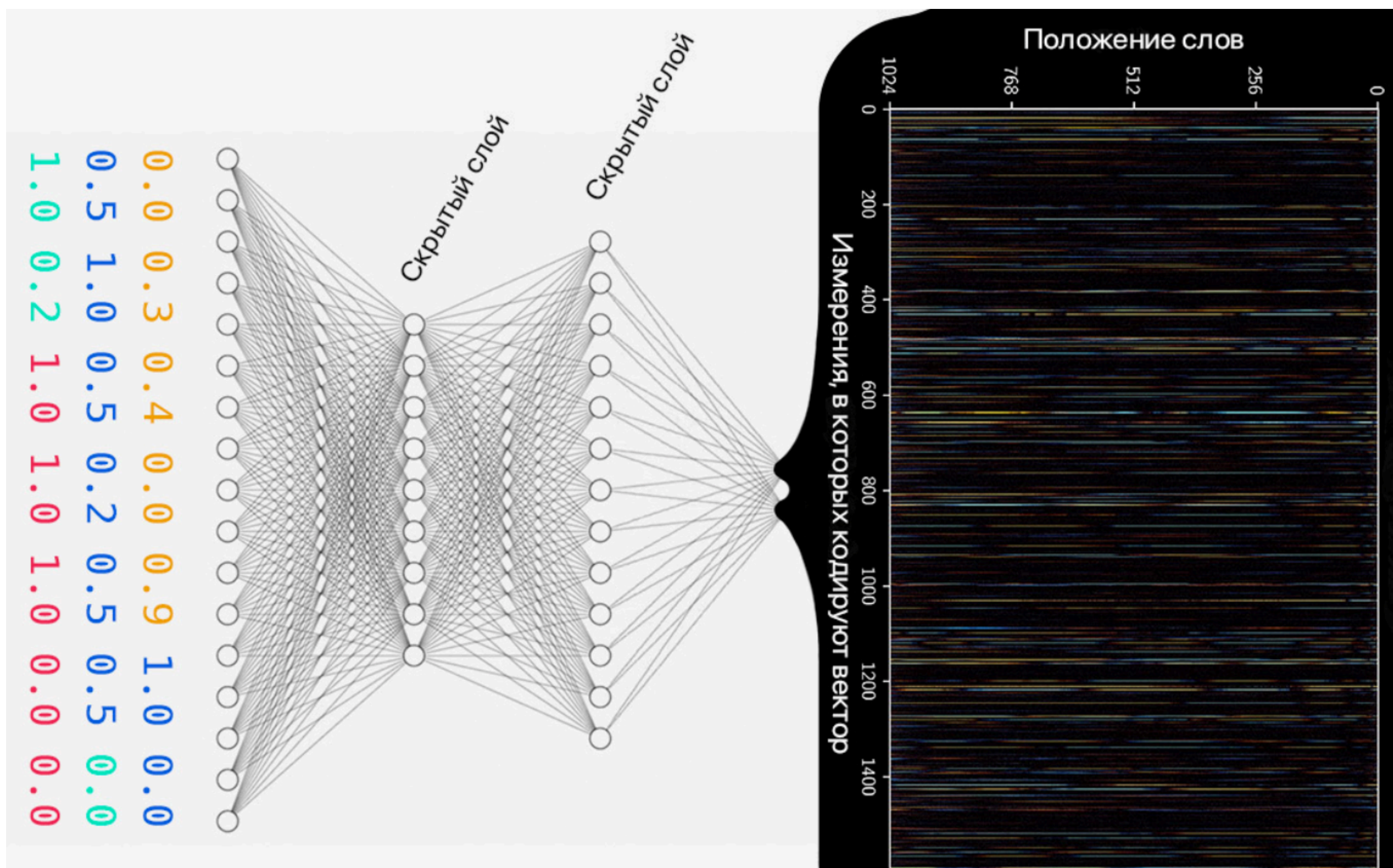
TEXT

TOKEN IDS

<https://platform.openai.com/tokenizer>

Как генерируют текст?

Токены в векторы



Как генерируют текст?

Кодирование векторов

Тут происходит тот самый черный ящик.

Основная технология ChatGPT - трансформер.

Задача: прогонять набор цифр (векторы) из прошлого шага через несколько слоёв кодирования.

Упрощённо говоря:

один слой разбирается, как каждое введённое слово соотносится с другим, чтобы понять ход мысли.

Другой слой генерирует тему обсуждения, на которой будет базироваться ответ.

Третий слой объединяет эти данные и соотносит их с разными *параметрами* или *измерениями* (категориями мышления).

Как генерируют текст?

Взвешивание

Задача на этом этапе оценить контекст и предсказать лучшую последовательность слов для ответа.

Токены:

[“Чинить”, “телефон”,
“не”, “включается”,
“после”, “вода”]

“Чинить” - 0.7

“телефон” - 0.9

“не”, “включается” - 0.3

“после” - 0.1

“вода” - 0.5

Как генерируют текст?

Декодирование и генерация текста.

Ответ формируется не одновременно.

Итоговый результат генерируется по одному слову за раз. После каждого слова уже написанная часть отправляется обратно для оценки, добавляя ещё один фактор в генератор. Уже написанный текст снова взвешивается, и учитывает вместе с факторами выше при создании следующего. И так до тех пор, пока бот не сгенерирует токен «стоп».

Как ГРТ понимает контекст?

Как GPT понимает контекст?

Объемы информации

Языковую модель тренировали на 45 ТБ текста.

Среди них: книги, статьи, страницы Википедии.

Сейчас. ChatGPT подключили к интернету, поэтому объем изученной информации - увеличился кратно!

Is GPT-generated?

Is GPT-generated?

Методы

- Большинство сервисов по определению авторства текстов основываются на:
 1. OpenAI Tokens
 2. Информация из интернета
 3. Машинное обучение самих сервисов

Сайты

Помогут распознать GPT-generated текст.

- <https://gptzero.me/>
- <https://www.zerogpt.com/>
- <https://contentatscale.ai/ai-content-detector/>
- <https://zerogpt.cc/>
- <https://x.writefull.com/gpt-detector>
- <https://writer.com/ai-content-detector/>
- <https://crossplag.com/ai-content-detector/>

Пример.

Курсовая выпускницы НИУ ВШЭ

Literature review

COVID-19 Impact

In view of the COVID-19 pandemic, a disease caused by a coronavirus infection, there was a crisis in the world and companies from different industries had to adopt new safety standards, reorganize work processes and incur losses. It is also worth adding about difficulties in managing the workforce within the company since HR managers have to create the most favorable atmosphere regardless of the current global problems (Carnevale & Hatak, 2020).

Considering the situation on a large-scale, the leaders of the companies had to transfer most of their employees to a remote work format, in connection with which the following series of problems occurred: lack of working equipment for work (laptop, printer, microphone for video meetings, software for getting the job done). The change in the format of work and adaptation to teleworking caused stress among a certain part of the population, thereby reducing the involvement and productivity of performing work tasks. Staff who have lost interest and focus at work tend to be aggressive and stressed, which ultimately will negatively affect the overall results of the company (Maslach & Leiter, 2016). Moreover, employee



Certainly, here's a literature review section for your course paper on the theme "Impact of the Remote Work Format on Employee Engagement During the COVID-19 Pandemic":

Literature Review

The COVID-19 pandemic, which emerged in late 2019, brought about unprecedented disruptions to global societies, including the way we work. As governments around the world implemented strict lockdown measures to curb the spread of the virus, organizations swiftly pivoted to remote work arrangements to ensure business continuity. Consequently, the remote work format became the new normal for millions of employees. This shift had a profound impact on employee engagement, raising questions about its long-term effects on productivity, well-being, and job satisfaction.

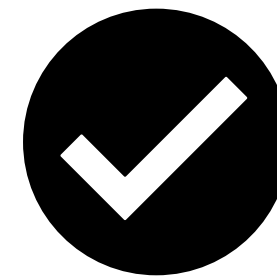
1. Remote Work and Employee Engagement

Remote work, often referred to as telecommuting or teleworking, involves employees working outside the traditional office setting, typically from their homes or other remote locations. The concept of remote work is not new, but the scale at which it was adopted during the COVID-19 pandemic was unprecedented.

Research conducted before the pandemic suggested mixed findings regarding the impact of remote work on employee engagement. Some studies argued that remote work could lead to increased engagement due to reduced commute times and greater autonomy. Others cautioned that it might result in feelings of isolation and decreased collaboration, potentially lowering engagement levels.

Пример.

Курсовая выпускницы НИУ ВШЭ



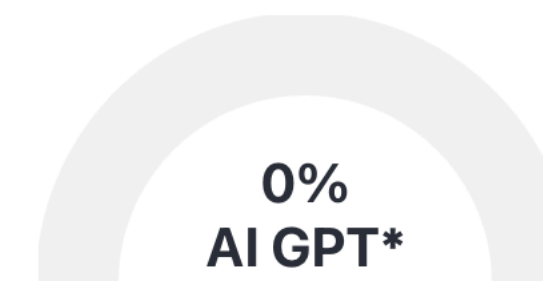
Literature review

COVID-19 Impact

In view of the COVID-19 pandemic, a disease caused by a coronavirus infection, there was a crisis in the world and companies from different industries had to adopt new safety standards, reorganize work processes and incur losses. It is also worth adding about difficulties in managing the workforce within the company since HR managers have to create the most favorable atmosphere regardless of the current global problems (Carnevale & Hatak, 2020).

Considering the situation on a large-scale, the leaders of the companies had to transfer most of their employees to a remote work format, in connection with which the following series of problems occurred: lack of working equipment for work (laptop, printer, microphone for video meetings, software for getting the job done). The change in the format of work and adaptation to teleworking caused stress among a certain part of the population, thereby reducing the involvement and productivity of performing work tasks. Staff who have lost interest and focus at work tend to be aggressive and stressed, which ultimately will negatively affect the overall results of the company (Maslach & Leiter, 2016). Moreover, employee

Your Text is Human written



HUMAN
Written

99.94%

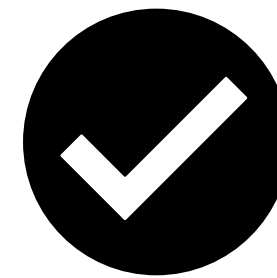
Predicted based upon **198 words**

AI/GPT
Generated

0.06%

Пример.

Курсовая выпускницы НИУ ВШЭ



Literature review

COVID-19 Impact

In view of the COVID-19 pandemic, a disease caused by coronavirus infection, there was a crisis in the world and companies from different industries had to adopt new safety standards, reorganize work processes and incur losses. It is also worth adding about difficulties in managing the workforce within the company since HR managers have to create the most favorable atmosphere regardless of the current global problems (Carnevale & Hatak, 2020).

Considering the situation on a large-scale, the leaders of the companies had to transfer most of their employees to a remote work format, in connection with which the following series of problems occurred: lack of working equipment for video meetings, software for getting the job done). The change in the format of work and adaptation to teleworking caused stress among a certain part of the population, thereby reducing the involvement and productivity of performing work tasks. Staff who have lost interest and focus at work tend to be aggressive and stressed, which ultimately will negatively affect the overall results of the company (Maslach & Leiter, 2016). Moreover, employee

Human Probability



Passes as Human!



Want Undetectable AI Content? Use Our AI Detector Pro!

Rewrite All AI Sentences

Our AI Detector pro tool uses a



B i A:

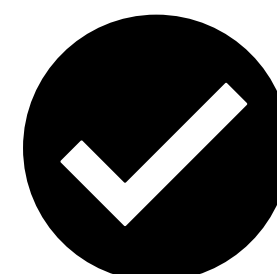


In view of the COVID-19 pandemic, a disease caused by a coronavirus infection, there was a crisis in the world and companies from different industries had to adopt new safety standards, reorganize work processes and incur losses. It is also worth adding about difficulties in managing the workforce within the company since HR managers have to create the most favorable atmosphere regardless of the current global problems (Carnevale & Hatak, 2020).

Considering the situation on a large-scale, the leaders of the companies had to transfer most of their employees to a remote work format, in connection with which the following series of problems occurred: lack of working equipment for work (laptop, printer, microphone for video meetings, software for getting the job done). The change in the format of work and adaptation to teleworking caused stress among a certain part of the population, thereby reducing the involvement and productivity of performing work tasks. Staff who have lost interest and focus at work tend to be aggressive and stressed, which ultimately will negatively affect the overall results of the company (Maslach & Leiter, 2016). Moreover, employee burnout has developed on the scale of a global problem (Gabriel & Aguinis, 2021).

Пример.

Курсовая выпускницы НИУ ВШЭ

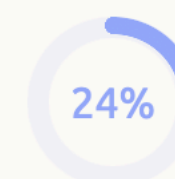


Literature review

COVID-19 Impact

In view of the COVID-19 pandemic, a disease caused by a coronavirus infection, there was a crisis in the world and companies from different industries had to adopt new safety standards, reorganize work processes and incur losses. It is also worth adding about difficulties in managing the workforce within the company since HR managers have to create the most favorable atmosphere regardless of the current global problems (Carnevale & Hatak, 2020).

Considering the situation on a large-scale, the leaders of the companies had to transfer most of their employees to a remote work format, in connection with which the following series of problems occurred: lack of working equipment for work (laptop, printer, microphone for video meetings, software for getting the job done). The change in the format of work and adaptation to teleworking caused stress among a certain part of the population, thereby reducing the involvement and productivity of performing work tasks. Staff who have lost interest and focus at work tend to be aggressive and stressed, which ultimately will negatively affect the overall results of the company (Maslach & Leiter, 2016). Moreover, employee



This text is likely to be written by **a human**
There is a **24%** probability this text was entirely written by AI

Sentences that are likely written by AI are **highlighted**.

In view of the COVID-19 pandemic, a disease caused by a coronavirus infection, there was a crisis in the world and companies from different industries had to adopt new safety standards, reorganize work processes and incur losses. It is also worth adding about difficulties in managing the workforce within the

Check ⚡

52% likely this comes from GPT-3, GPT-4 or ChatGPT.

Пример.

Курсовая выпускницы НИУ ВШЭ



Certainly, here's a literature review section for your course paper on the theme "Impact of the Remote Work Format on Employee Engagement During the COVID-19 Pandemic":

Literature Review

The COVID-19 pandemic, which emerged in late 2019, brought about unprecedented disruptions to global societies, including the way we work. As governments around the world implemented strict lockdown measures to curb the spread of the virus, organizations swiftly pivoted to remote work arrangements to ensure business continuity. Consequently, the remote work format became the new normal for millions of employees. This shift had a profound impact on employee engagement, raising questions about its long-term effects on productivity, well-being, and job satisfaction.

1. Remote Work and Employee Engagement

Remote work, often referred to as telecommuting or teleworking, involves employees working outside the traditional office setting, typically from their homes or other remote locations. The concept of remote work is not new, but the scale at which it was adopted during the COVID-19 pandemic was unprecedented.

Research conducted before the pandemic suggested mixed findings regarding the impact of remote work on employee engagement. Some studies argued that remote work could lead to increased engagement due to reduced commute times and greater autonomy. Others cautioned that it might result in feelings of isolation and decreased collaboration, potentially lowering engagement levels.

↻ Regen



Hello! Write Literature review of COVID-19 Impact in a course paper on the theme Impact of the Remote Work Format on Employee Engagement. During the COVID-19 Pandemic

Пример.

Курсовая выпускницы НИУ ВШЭ



Certainly, here's a literature review section for your course paper on the theme "Impact of the Remote Work Format on Employee Engagement During the COVID-19 Pandemic":

Literature Review

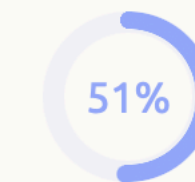
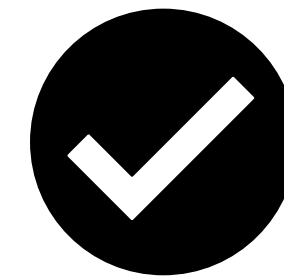
The COVID-19 pandemic, which emerged in late 2019, brought about unprecedented disruptions to global societies, including the way we work. As governments around the world implemented strict lockdown measures to curb the spread of the virus, organizations swiftly pivoted to remote work arrangements to ensure business continuity. Consequently, the remote work format became the new normal for millions of employees. This shift had a profound impact on employee engagement, raising questions about its long-term effects on productivity, well-being, and job satisfaction.

1. Remote Work and Employee Engagement

Remote work, often referred to as telecommuting or teleworking, involves employees working outside the traditional office setting, typically from their homes or other remote locations. The concept of remote work is not new, but the scale at which it was adopted during the COVID-19 pandemic was unprecedented.

Research conducted before the pandemic suggested mixed findings regarding the impact of remote work on employee engagement. Some studies argued that remote work could lead to increased engagement due to reduced commute times and greater autonomy. Others cautioned that it might result in feelings of isolation and decreased collaboration, potentially lowering engagement levels.

↻ Regen



This text is moderately likely to be written AI
There is a 51% probability this text was entirely written by AI

Sentences that are likely written by AI are highlighted.

questions about its long-term effects on productivity, well-being, and job satisfaction.

1. Remote Work and Employee Engagement

Remote work, often referred to as telecommuting or teleworking, involves employees working outside the traditional office setting, typically from their homes or other remote locations. The concept of remote work is not new, but the scale at which it was adopted during the COVID-19 pandemic was unprecedented.



Your Text is Most Like

The COVID-19 pandemic, which emerged in late 2019, brought about unprecedented disruptions to global societies, including the way we work. As governments around the world implemented strict lockdown measures to curb the spread of the virus, organizations swiftly pivoted to remote work arrangements to ensure business continuity. Consequently, the remote work format became the new normal for millions of employees. This shift had a profound impact on employee engagement, raising questions about its long-term effects on productivity, well-being, and job satisfaction.

1. Remote Work and Employee Engagement

Remote work, often referred to as telecommuting or teleworking, involves employees working outside the traditional office setting, typically from their homes or other remote locations. The concept of remote work is not new, but the scale at which it was adopted during the COVID-19 pandemic was unprecedented.

Пример.

Курсовая выпускницы НИУ ВШЭ



Certainly, here's a literature review section for your course paper on the theme "Impact of the Remote Work Format on Employee Engagement During the COVID-19 Pandemic":

Literature Review

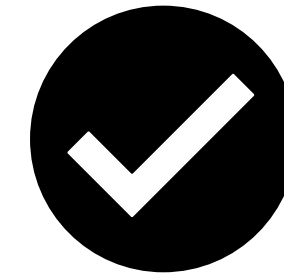
The COVID-19 pandemic, which emerged in late 2019, brought about unprecedented disruptions to global societies, including the way we work. As governments around the world implemented strict lockdown measures to curb the spread of the virus, organizations swiftly pivoted to remote work arrangements to ensure business continuity. Consequently, the remote work format became the new normal for millions of employees. This shift had a profound impact on employee engagement, raising questions about its long-term effects on productivity, well-being, and job satisfaction.

1. Remote Work and Employee Engagement

Remote work, often referred to as telecommuting or teleworking, involves employees working outside the traditional office setting, typically from their homes or other remote locations. The concept of remote work is not new, but the scale at which it was adopted during the COVID-19 pandemic was unprecedented.

Research conducted before the pandemic suggested mixed findings regarding the impact of remote work on employee engagement. Some studies argued that remote work could lead to increased engagement due to reduced commute times and greater autonomy. Others cautioned that it might result in feelings of isolation and decreased collaboration, potentially lowering engagement levels.

↻ Regen



Check ⚡

1% likely this comes from GPT-3, GPT-4 or ChatGPT.

HUMAN
Written

99.98%

Predicted based upon **187 words**

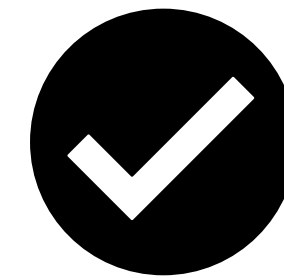
AI/GPT
Generated

0.02%

Highly likely to be Human!

Пример.

Курсовая выпускницы НИУ ВШЭ



Certainly, here's a literature review section for your course paper on the theme "Impact of the Remote Work Format on Employee Engagement During the COVID-19 Pandemic":

Literature Review

The COVID-19 pandemic, which emerged in late 2019, brought significant disruptions to global societies, including the way we work. As governments around the world implemented strict lockdown measures to curb the spread of the virus, many organizations swiftly pivoted to remote work arrangements to ensure business continuity. Consequently, the remote work format became the new normal for millions of employees. This shift had a profound impact on employee engagement, raising questions about its long-term effects on productivity, well-being, and job satisfaction.

1. Remote Work and Employee Engagement

Remote work, often referred to as telecommuting or teleworking, involves employees working outside the traditional office setting, typically from their homes or other remote locations. The concept of remote work is not new, but the scale at which it was adopted during the COVID-19 pandemic was unprecedented.

Research conducted before the pandemic suggested mixed findings regarding the impact of remote work on employee engagement. Some studies argued that remote work could lead to increased engagement due to reduced commute times and greater autonomy. Others cautioned that it might result in feelings of isolation and decreased collaboration, potentially lowering engagement levels.

Human Probability


Reads like AI! 😞

Want Undetectable AI Content? Use Our AI Detector Pro!

Rewrite All AI Sentences

Our AI Detector pro tool uses a proprietary mix of 3 AI engines, NLP and semantic analysis algorithms to rewrite sentences and paraphrase paragraphs in a



B i A:

This reads very robotic and therefore has a higher

The COVID-19 pandemic, which emerged in late 2019, brought significant disruptions to global societies, including the way we work. As governments around the world implemented strict lockdown measures to curb the spread of the virus, many organizations swiftly pivoted to remote work arrangements to ensure business continuity. Consequently, the remote work format became the new normal for millions of employees. This shift had a profound impact on employee engagement, raising questions about its long-term effects on productivity, well-being, and job satisfaction.

1. Remote Work and Employee Engagement

Remote work, often referred to as telecommuting or teleworking, involves employees working outside the traditional office setting, typically from their homes or other remote locations. The concept of remote work is not new, but the scale at which it was adopted during the COVID-19 pandemic was unprecedented.

Research conducted before the pandemic suggested mixed findings regarding the impact of remote work on employee engagement. Some studies argued that remote work could lead to increased engagement due to reduced commute times and greater autonomy. Others cautioned that it might result in feelings of isolation and decreased collaboration, potentially lowering engagement levels.

Regen

Пример.

Курсовая выпускницы НИУ ВШЭ



Certainly, here's a literature review section for your course paper on the theme "Impact of the Remote Work Format on Employee Engagement During the COVID-19 Pandemic":

Literature Review

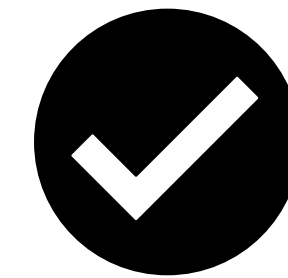
The COVID-19 pandemic, which emerged in late 2019, brought about unprecedented disruptions to global societies, including the way we work. As governments around the world implemented strict lockdown measures to curb the spread of the virus, organizations swiftly pivoted to remote work arrangements to ensure business continuity. Consequently, the remote work format became the new normal for millions of employees. This shift had a profound impact on employee engagement, raising questions about its long-term effects on productivity, well-being, and job satisfaction.

1. Remote Work and Employee Engagement

Remote work, often referred to as telecommuting or teleworking, involves employees working outside the traditional office setting, typically from their homes or other remote locations. The concept of remote work is not new, but the scale at which it was adopted during the COVID-19 pandemic was unprecedented.

Research conducted before the pandemic suggested mixed findings regarding the impact of remote work on employee engagement. Some studies argued that remote work could lead to increased engagement due to reduced commute times and greater autonomy. Others cautioned that it might result in feelings of isolation and decreased collaboration, potentially lowering engagement levels.

↻ Regen



Ask chatGPT!



No, I did not write that specific passage. It appears to be a description or analysis of the impact of the COVID-19 pandemic on the way people work and its effects on employee engagement, productivity, well-being, and job satisfaction. If you have any questions or would like to discuss this topic further, please feel free to ask, and I'd be happy to provide more information or insights.

Дискуссия.

Как вы считаете, какие есть способы распознавания студенческих работ на генерацию Chat GPT?

References

- <https://zapier.com/blog/how-does-chatgpt-work/>
- <https://www.techtarget.com/whatis/definition/ChatGPT#:~:text=ChatGPT uses deep learning, a,its training data's typical sequence.>
- <https://gilberttanner.com/blog/generating-text-using-a-recurrent-neuralnetwork/>
- <https://towardsdatascience.com/training-neural-networks-to-create-text-like-a-human-23bfdc23c28>
- <https://openai.com/>
- <https://techwiser.com/ways-to-check-if-the-text-is-ai-generated-by-chatgpt/>

Благодарю за внимание!
Готов ответить на ваши вопросы.

Александр Богданов
Научно-учебная группа по исследованию частотных характеристик языка
НИУ ВШЭ – Санкт-Петербург
10.10.2023