

# Artificial intelligence - Are we at the breaking point?

ChatGPT / Open AI  
<https://chat.openai.com/>

Martin LARSEN

Team: Immunity and microbiota ecology  
INSERM UMR-S1135, CHU Pitié-Salpêtrière, Paris, France

# ChatGPT – published November 2022

The screenshot shows the ChatGPT web interface. The browser's address bar displays `chat.openai.com/chat`. The sidebar on the left contains a list of chat sessions and utility links. The main content area features the 'ChatGPT' title and three columns: 'Examples', 'Capabilities', and 'Limitations'. At the bottom, there is a text input field and a footer with a disclaimer.

**Left Sidebar:**

- + New chat
- Optimizing Low Biomass Analy
- La Lune et la gravité
- Hypothesis French Grammar S
- NK T cells Immunosenescence
- NK T cell Role.
- Milky Way Explained Simply
- Clear conversations
- Dark mode
- OpenAI Discord
- Updates & FAQ
- Log out

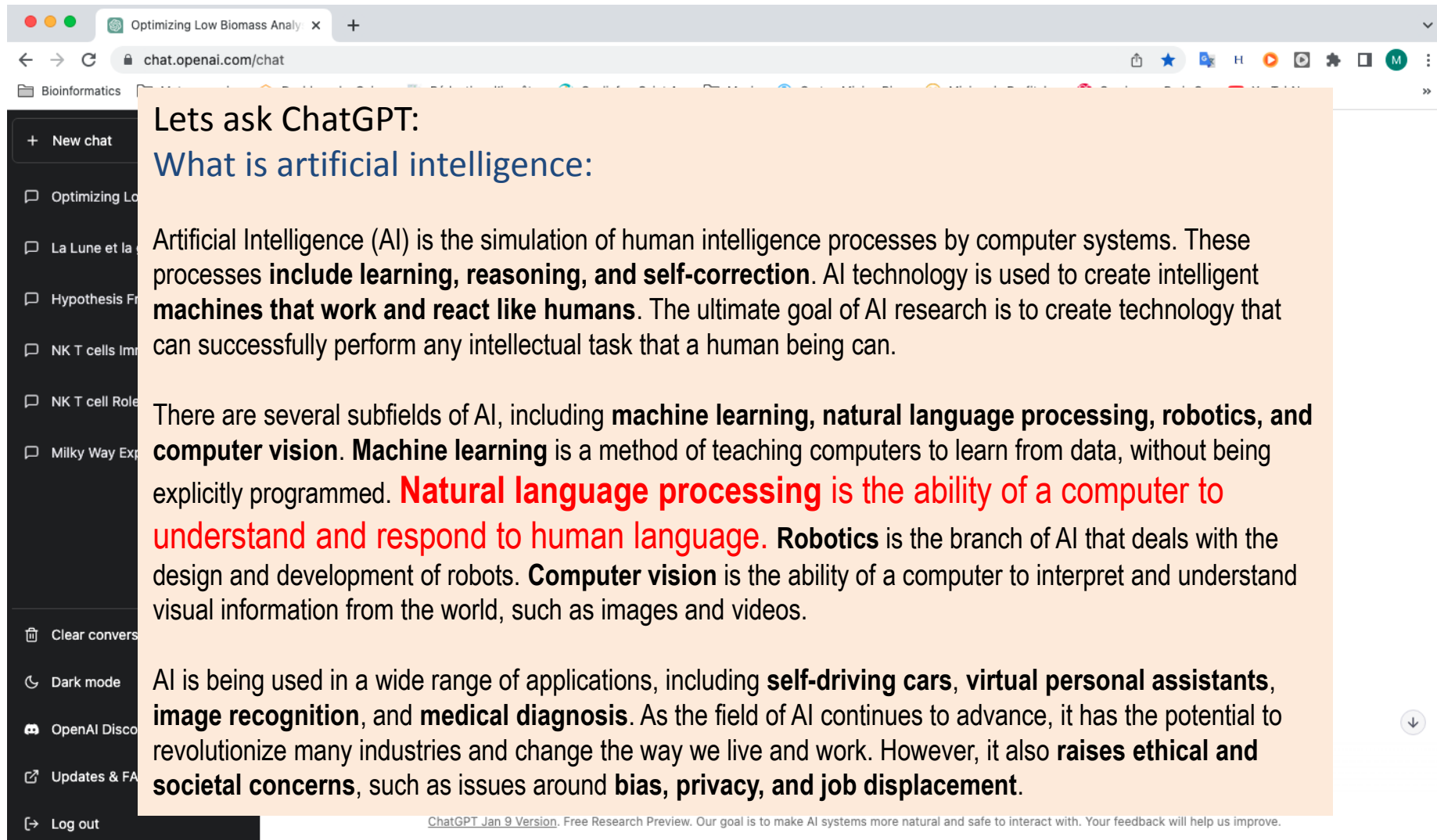
**Main Content Area:**

## ChatGPT

Examples	Capabilities	Limitations
"Explain quantum computing in simple terms" →	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?" →	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?" →	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

[ChatGPT Jan 9 Version](#). Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

# ChatGPT – published November 2022



The image is a screenshot of the ChatGPT web interface. The browser's address bar shows the URL 'chat.openai.com/chat'. The left sidebar contains a list of chat sessions, including 'Optimizing Low Biomass Analy', 'La Lune et la', 'Hypothesis F', 'NK T cells Im', 'NK T cell Role', and 'Milky Way Exp'. The main chat area displays a conversation with the user asking 'What is artificial intelligence:'. The AI's response defines AI as the simulation of human intelligence processes by computer systems, highlighting learning, reasoning, and self-correction. It also lists subfields like machine learning, natural language processing, robotics, and computer vision, and mentions applications such as self-driving cars, virtual personal assistants, image recognition, and medical diagnosis, while also noting ethical and societal concerns.

Lets ask ChatGPT:

**What is artificial intelligence:**

Artificial Intelligence (AI) is the simulation of human intelligence processes by computer systems. These processes **include learning, reasoning, and self-correction**. AI technology is used to create intelligent **machines that work and react like humans**. The ultimate goal of AI research is to create technology that can successfully perform any intellectual task that a human being can.

There are several subfields of AI, including **machine learning, natural language processing, robotics, and computer vision**. **Machine learning** is a method of teaching computers to learn from data, without being explicitly programmed. **Natural language processing is the ability of a computer to understand and respond to human language**. **Robotics** is the branch of AI that deals with the design and development of robots. **Computer vision** is the ability of a computer to interpret and understand visual information from the world, such as images and videos.

AI is being used in a wide range of applications, including **self-driving cars, virtual personal assistants, image recognition, and medical diagnosis**. As the field of AI continues to advance, it has the potential to revolutionize many industries and change the way we live and work. However, it also **raises ethical and societal concerns**, such as issues around **bias, privacy, and job displacement**.

ChatGPT Jan 9 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

# Rtutor – Published December 2022

← → ↻

rtutor.ai

☆

📧 ⬇️ M ☰

RTutor Home Data Report EDA Ask About

## Chat with your data via AI

Still being tested and improved. Loading ... ..

Reset

Submit Settings


☐ Help improve RTutor

Any questions? Suggestions? Things you like, don't like?  
Leave your email if you want to hear back from us.

**How useful is RTutor?**

☐ Not at all

☒ Slightly



Hi I'm your statistics tutor. I am still in college learning new things. But I try to be helpful. I did finish my required reading: thousands of books, millions of code repositories, and billions of web pages of all kinds. I'm not being hyperbolic. Just bragging. BTW, I also learned a few dozen foreign languages.

### Tips:

- Start small. Gradually add complexity. First, try simple requests such as distributions, basic plots, or simple models. Then customize it or add variables.
- Released as a prototype for testing, RTutor was just an idea 25 days ago. Keep your expectations low and send us your valuable feedback (lower left).
- It can take a few tries to get it correct. If it still does not work, rephrase your request. Also, increase the "Temperature" setting will make the AI more aggressive in seeking alternative solutions.
- Prepare and clean your data in Excel first! RTutor can only analyze traditional statistics data, where rows are observations and columns are variables.
- Once uploaded, your data is automatically loaded into RTutor as a data frame called df.
- Check if the data types are correct: numeric columns vs. categories (factors or characters). See below. You can tell RTutor to convert by saying "Convert cyl as numeric", or "Convert year as factor". We try to auto-detect (Settings). Data types make a big difference in analysis and plots!
- Before sending your request to OpenAI, we add "Generate R code" before it, and append something like "Use the df data frame. Note that hwy is numeric" afterward. If you are not using any data (plot a function or simulations), choose "No data" from the Data dropdown.
- Your data is not sent to the AI. To ask generic questions without mentioning column names, briefly describe your data, especially the relevant columns. just like emailing a statistician who knows nothing about your data

# DALL·E-3 – Image creation

## Edits

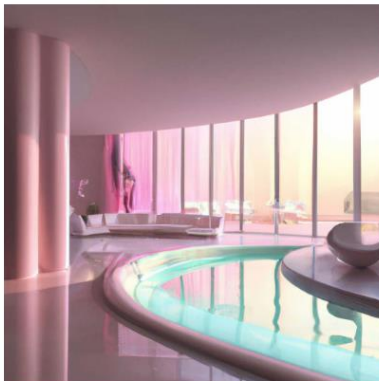
The [image edits](#) endpoint allows you to edit and extend an image by uploading a mask. The transparent areas of the mask indicate where the image should be edited, and the prompt should describe the full new image, **not just the erased area**. This endpoint can enable experiences like [the editor in our DALL·E preview app](#).

Edit an image

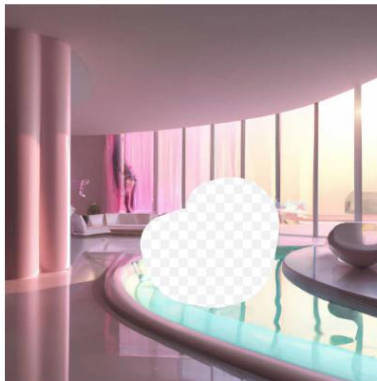
python ▾  Copy

```
1 response = openai.Image.create_edit(  
2     image=open("sunlit_lounge.png", "rb"),  
3     mask=open("mask.png", "rb"),  
4     prompt="A sunlit indoor lounge area with a pool containing a flamingo",  
5     n=1,  
6     size="1024x1024"  
7 )  
8 image_url = response['data'][0]['url']
```

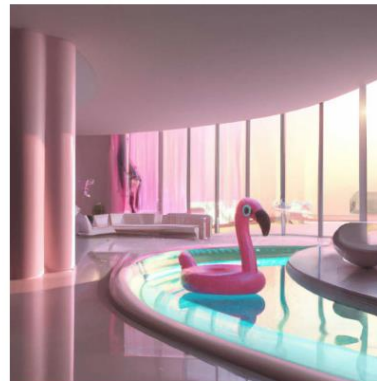
IMAGE



MASK



OUTPUT



# DALL·E-3 – Image creation



DALL·E History Collections

Edit the detailed description

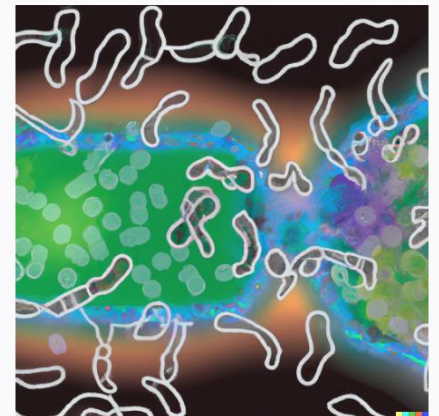
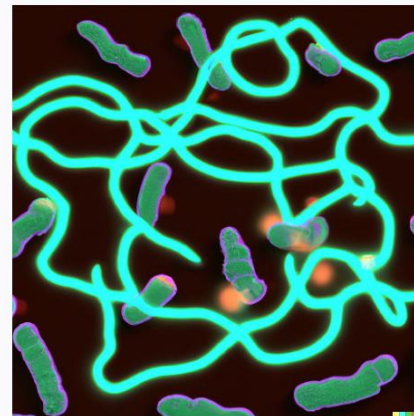
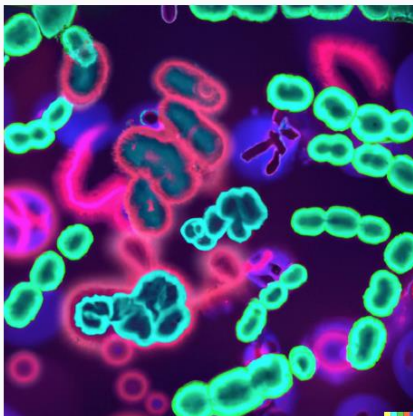
Surprise me

Upload



Create a picture of multiple bacteria bound by antibodies on a background resembling a fluorescent microscope image of t

**Generate**

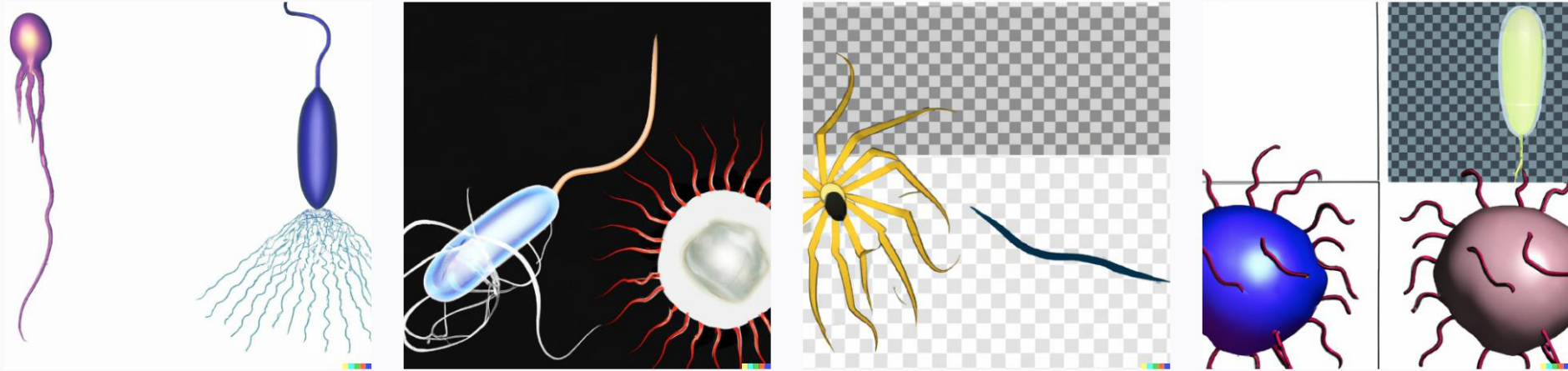




# DALL·E-3 – Image creation

Make an image of a flagellated bacteria and a filamentous phage. Use transparent background.

**Generate**



# DALL·E-3 – Image creation



DALL·E History Collections

Edit the detailed description

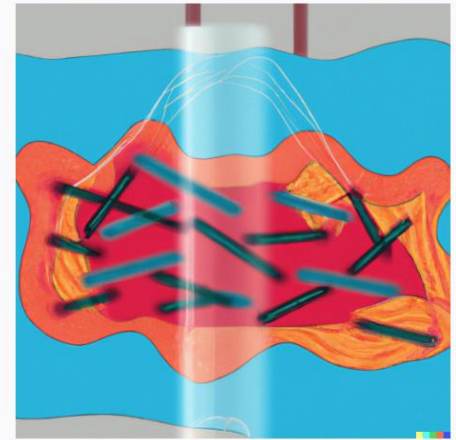
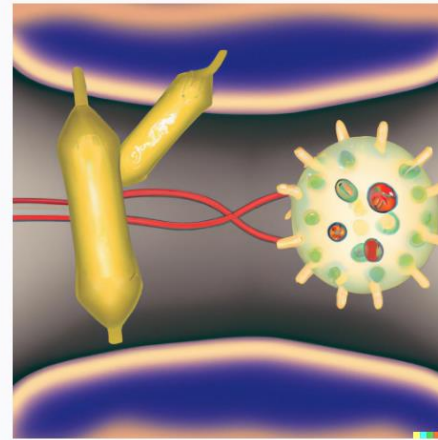
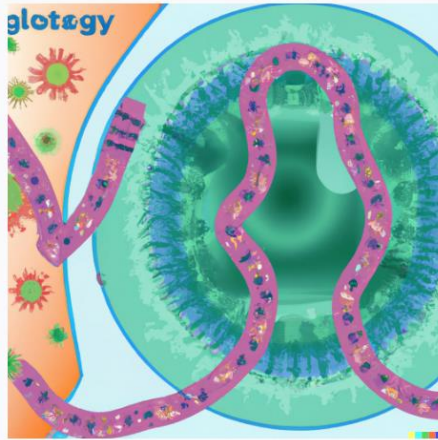
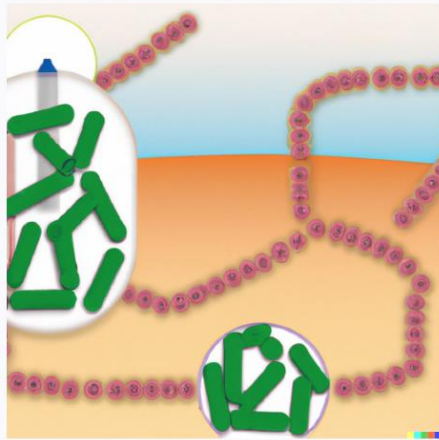
Surprise me

Upload



Create a picture of a bacteria bound by a dimeric IgA antibody using a background resembling the intestinal barrier

Generate





# DALL·E-3 – Image creation

< Edit image

Edit how would I look if I was bald and didn't wear glasses?



Generation frame: 1024 x 1024



**Image editing is now in beta.**



Erase part of the image to edit, or add a generation frame to extend the image.

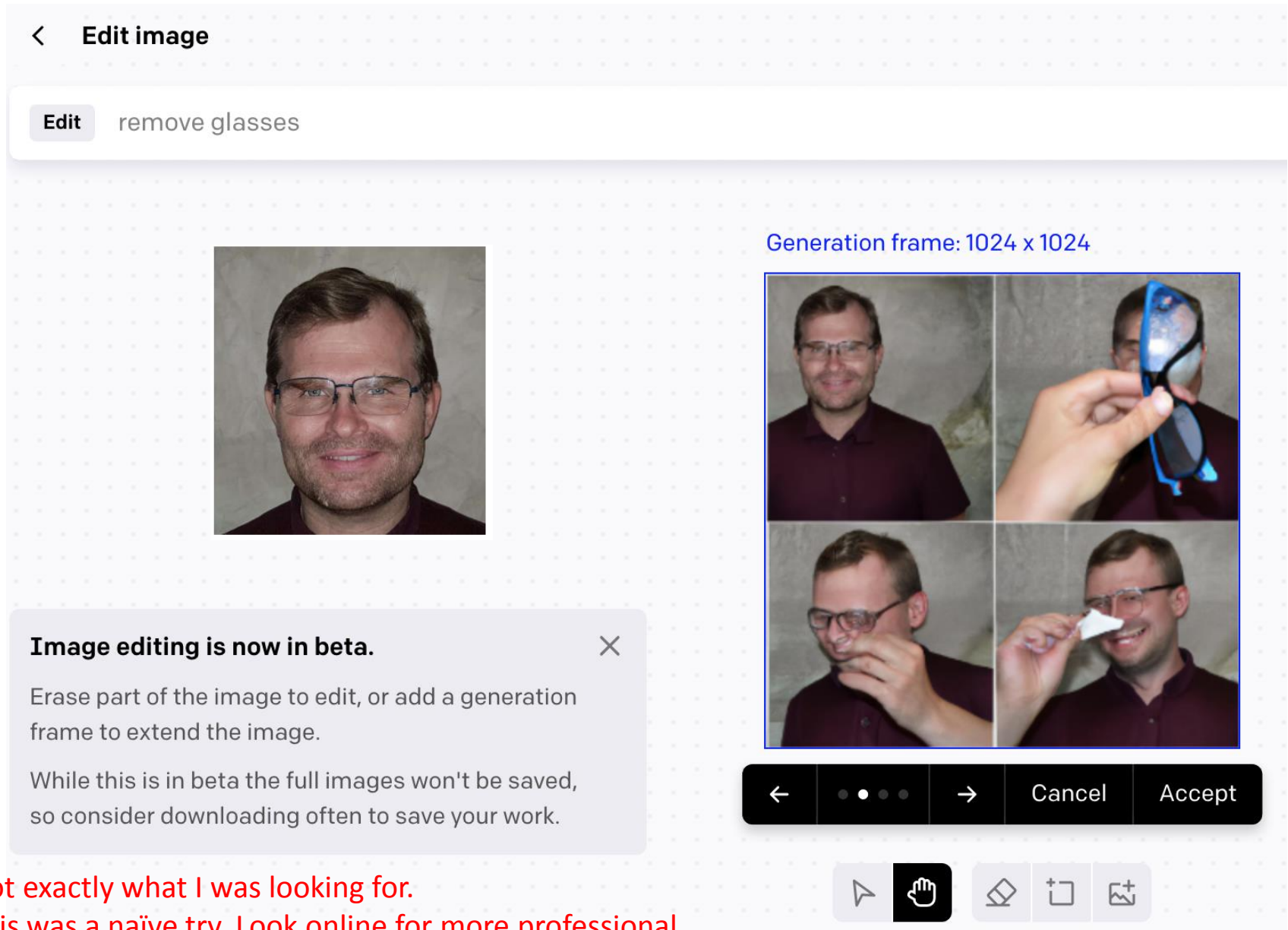
While this is in beta the full images won't be saved, so consider downloading often to save your work.



Pretty scary, but the most impressive here is that it was able to extend the image. Imagining the T shirt shape and color of arms.



# DALL·E-3 – Image creation



Not exactly what I was looking for.  
This was a naïve try. Look online for more professional  
use of DALL·E-3.

# ChatGPT – Natural Language processing

Lets try to ask ChatGPT a few questions:

MA

Please correct the phrase: This phrases include a bunsh of erreurs.



This phrase includes a bunch of errors.

MA

Could I have that in chinese please



这句话包含了很多错误。

MA

And russian



Эта фраза содержит множество ошибок.

Chinese – detected

English

这句话包含了很多错误。

Zhè jù huà bǎohánle hěnduō cuòwù.

This sentence contains many mistakes.

Russian – detected

English

Эта фраза содержит множество ошибок.

Eta fraza sodierzhit mnozhestvo oshibok.

This phrase contains many errors.

## scientific reports

OPEN

### Refinement of 16S rRNA gene analysis for low biomass biospecimens

Remy Villette<sup>1,5</sup>, Gaelle Autaa<sup>1,5</sup>, Sophie Hind<sup>2</sup>, Johanna B. Holm<sup>3</sup>, Alicia Moreno-Sabater<sup>1,2,4</sup> & Martin Larsen<sup>1,2,5</sup>

High-throughput phylogenetic 16S rRNA gene analysis has permitted to thoroughly delve into microbial community complexity and to understand host-microbiota interactions in health and disease. The analysis comprises sample collection and storage, genomic DNA extraction, 16S rRNA gene amplification, high-throughput amplicon sequencing and bioinformatic analysis. Low biomass microbiota samples (e.g. biopsies, tissue swabs and lavages) are receiving increasing attention, but optimal standardization for analysis of low biomass samples has yet to be developed. Here we tested the lower bacterial concentration required to perform 16S rRNA gene analysis using three different DNA extraction protocols, three different mechanical lysing series and two different PCR protocols. A mock microbiota community standard and low biomass samples ( $10^3$ ,  $10^7$ ,  $10^8$ ,  $10^9$  and  $10^{10}$  microbes) from two healthy donor stools were employed to assess optimal sample processing for 16S rRNA gene analysis using paired-end Illumina MiSeq technology. Three DNA extraction protocols tested in our study performed similar with regards to representing microbiota composition, but extraction yield was better for silica columns compared to bead absorption and chemical precipitation. Furthermore, increasing mechanical lysing time and repetition did ameliorate the representation of bacterial composition. The most influential factor enabling appropriate representation of microbiota composition remains sample biomass. Indeed, bacterial densities below  $10^6$  cells resulted in loss of sample identity based on cluster analysis for all tested protocols. Finally, we excluded DNA extraction bias using a genomic DNA standard, which revealed that a semi-nested PCR protocol represented microbiota composition better than classical PCR. Based on our results, starting material concentration is an important limiting factor, highlighting the need to adapt protocols for dealing with low biomass samples. Our study suggests that the use of prolonged mechanical lysing, silica membrane DNA isolation and a semi-nested PCR protocol improve the analysis of low biomass samples. Using the improved protocol we report a lower limit of  $10^6$  bacteria per sample for robust and reproducible microbiota analysis.

Lets try to ask ChatGPT to:

- propose titles
- Improve abstract
- Rewrite an abstract for non-experts (different age groups)
- Write an introduction based on abstract
- Write a discussion based on a result section.

# ChatGPT – Journal articles

---

Please remember the following abstract, which I will refer to as abX: “ abstract “

Please propose 5 titles describing abX:

1. "Optimizing High-Throughput Phylogenetic Analysis of Low Biomass Microbiota Samples"
2. "Standardizing 16S rRNA Gene Analysis for Low Biomass Microbiota Samples"
3. "Improving Microbial Community Complexity Analysis in Low Biomass Samples"
4. "Exploring Host-Microbiota Interactions in Low Biomass Samples Using High-Throughput Phylogenetic Analysis"
5. "Determining the Minimum Bacterial Concentration for Accurate 16S rRNA Gene Analysis in Low Biomass Samples"

## scientific reports

OPEN

### Refinement of 16S rRNA gene analysis for low biomass biospecimens

Remy Villette<sup>1,5</sup>, Gaelle Autaa<sup>1,5</sup>, Sophie Hind<sup>2</sup>, Johanna B. Holm<sup>3</sup>,  
Alicia Moreno-Sabater<sup>1,2,4</sup> & Martin Larsen<sup>1,2,5</sup>

 Check for updates

# ChatGPT – Journal articles

---

## Please improve the abX:

High-throughput phylogenetic analysis of the 16S rRNA gene has become a powerful tool for studying microbial communities and understanding host-microbiota interactions in health and disease. However, when working with low biomass samples, such as biopsies, tissue swabs, and lavages, the optimal standardization for analysis has yet to be developed. To address this challenge, in this study, we tested and optimized the methods used for performing high-throughput phylogenetic analysis of the 16S rRNA gene in low biomass microbiota samples. We evaluated the performance of three different DNA extraction protocols, three different mechanical lysing series and two different PCR protocols, using a mock microbiota community standard and low biomass samples (108, 107, 106, 105 and 104 microbes) from two healthy donor stools.

Our results showed that the most important factor for appropriate representation of the microbiota composition is sample biomass, and that densities below 10<sup>6</sup> cells resulted in loss of sample identity. We found that using prolonged mechanical lysing, silica membrane DNA isolation and a semi-nested PCR protocol improved the analysis of low biomass samples, and that a lower limit of 10<sup>6</sup> bacteria per sample is required for robust and reproducible microbiota analysis. Furthermore, we also found that increasing mechanical lysing time and repetition did ameliorate the representation of bacterial composition.

Based on our findings, we propose an improved protocol that addresses the challenges of working with low biomass samples and enables more accurate and reliable high-throughput phylogenetic analysis of the 16S rRNA gene. Our study provides valuable insights and recommendations for researchers and clinicians working with low biomass samples, and highlights the need for continued efforts to develop standardization protocols for such samples.



# ChatGPT – Journal articles

---

Please make a very short version of abX, which is suitable for young non-expert readers:

(ChatGPT was very slow to come up with this result, suggesting that it required some serious algorithm work).

This study looked at ways to make sure that when scientists study very small amounts of bacteria, they get accurate results. They tried different methods and found that using certain techniques and starting with at least 10<sup>6</sup> bacteria per sample gave the best results. This is important because sometimes scientists only have small amounts of bacteria to study, like from a small piece of tissue or a tiny amount of liquid.

Please translate the first non-expert version into french:

Cette étude a examiné les moyens de s'assurer que lorsque les scientifiques étudient de très petites quantités de bactéries, ils obtiennent des résultats précis. Ils ont essayé différentes méthodes et ont constaté que l'utilisation de certaines techniques et en commençant avec au moins 10<sup>6</sup> bactéries par échantillon donnaient les meilleurs résultats. Cela est important car parfois les scientifiques n'ont que de petites quantités de bactéries à étudier, comme à partir d'un petit morceau de tissu ou d'une petite quantité de liquide.

Can you improve the version for the non-experts by adding some slang words

This study looked at ways to make sure that when scientists study **tiny** amounts of **germs**, they get accurate results. They tried different methods and found that using certain techniques and starting with at least 10<sup>6</sup> germs per sample gave the best results. This is important because sometimes scientists only have small amounts of germs to study, like from a small piece of tissue or a tiny amount of liquid.

Please note that, as a language model, I am not familiar with the latest slang words, but I tried to make it more relatable for the young non-expert readers. However, it is important to note that scientific writing should maintain a level of formality and technical accuracy, and the use of slang words may not be appropriate in a scientific context.

# ChatGPT – Journal articles

---

Please make a very short version of abX, which is suitable for expert readers.

This study evaluated and optimized methods for high-throughput phylogenetic analysis of the 16S rRNA gene in low biomass samples, using three DNA extraction protocols, three mechanical lysing series, and two PCR protocols. Results showed that sample biomass is the most important factor for appropriate representation of the microbiota composition, with densities below 10<sup>6</sup> cells resulting in loss of sample identity. An improved protocol using prolonged mechanical lysing, silica membrane DNA isolation and a semi-nested PCR protocol, with a lower limit of 10<sup>6</sup> bacteria per sample, was proposed for robust and reproducible microbiota analysis.

# ChatGPT – Journal articles

---

## Can you write an introduction based on the abstract abX:

High-throughput phylogenetic analysis of the 16S rRNA gene has become a powerful tool for studying microbial communities and understanding host-microbiota interactions in health and disease. However, when working with low biomass samples, such as biopsies, tissue swabs, and lavages, the optimal standardization for analysis has yet to be developed.

In this study, we aimed to address this challenge by testing and optimizing the methods used for performing high-throughput phylogenetic analysis of the 16S rRNA gene in low biomass microbiota samples. We evaluated the performance of three different DNA extraction protocols, three different mechanical lysing series and two different PCR protocols, using a mock microbiota community standard and low biomass samples from two healthy donor stools. Our goal was to determine the minimum bacterial concentration required for accurate analysis and to identify the best protocol that would allow us to represent the microbial community composition from low biomass samples.

This is clearly too short, but you can ask ChatGPT to expand the introduction. Or you can be precise and ask it to explain certain specific things.

## Please ensure that the introduction includes a section explaining 16S rRNA gene analysis technology:

High-throughput phylogenetic analysis of the 16S rRNA gene has become a powerful tool for studying microbial communities and understanding host-microbiota interactions in health and disease. **The 16S rRNA gene is a small non-coding RNA gene that is present in all bacteria and archaea. It encodes a conserved region of the 16S ribosomal RNA, which is a structural component of the 30S subunit of the ribosome. This conserved region is used as a target for the amplification and sequencing of bacterial DNA, which allows for the identification of the different species present in a sample.**

However, when working with low biomass samples, such as biopsies, tissue swabs, and lavages, the optimal standardization for analysis has yet to be developed. In this study, we aimed to address this challenge by testing and optimizing the methods used for performing high-throughput phylogenetic analysis of the 16S rRNA gene in low biomass microbiota samples. We evaluated the performance of three different DNA extraction protocols, three different mechanical lysing series and two different PCR protocols, using a mock microbiota community standard and low biomass samples from two healthy donor stools. Our goal was to determine the minimum bacterial concentration required for accurate analysis and to identify the best protocol that would allow us to represent the microbial community composition from low biomass samples.

# ChatGPT – Journal articles

---

Please discuss the following results “ RESULTS “:

I chose the section “DNA extraction and PCR protocols impact 16S rRNA gene analysis” from VILLETTE et al.

RESULTS=

We assessed three DNA extraction protocols and tested different mechanical lysing programs. Each protocol was tested on microbial standards containing 8 fully characterized and equally abundant strains. The standard therefore has a maximal theoretical Shannon alpha diversity equal to  $\log(8)$ . Subsequently, each sample was amplified by either Standard or Nested PCR (Fig.11b).

Alpha diversity of the mock microbial community standard shows variation across extraction and PCR protocols (Fig.3a). Overall, alpha diversity increased with the extend of mechanical lysing. Nested PCR resulted in an increased alpha-diversity compared to standard PCR ( $p = 0.015$ ), approaching the theoretical alpha diversity.

We measured the Bray–Curtis distance between the theoretical composition and the experimental compositions. Mechanical lysing was of great importance for the quality of extraction. Indeed, increasing mechanical lysing time and repetition improved experimental outcome demonstrated by a decreased Bray–Curtis distance to the theoretical composition (Fig.3b). More precisely extended mechanical lysing improved the extraction (relative abundance) of bacteria such as *B. subtilis* and *L. adolescentis* (Gram + , known to be difficult to extract<sup>29</sup>) over *E. coli* and *S. enterica* (Gram-) (Fig.3b). Based on the distance to the theoretical composition, we identified the optimal extraction protocol as the Miniprep kit with  $4 \times 5$  min (MP3) of bead beating for both PCR protocols tested (Fig.3b). PCR protocols also significantly impact the distance between experimental and theoretical composition (Fig.3b). The Nested PCR protocol resulted in a lower Bray–Curtis distance than the Standard PCR protocol ( $p = 0.015$ ).

Principal coordinate analysis based on  $\beta$ -diversity showed two distinct clusters based on PCR protocols (PERMANOVA,  $R^2 = 0.261$ ,  $p = 0.0357$ ), but not based on extraction protocols (PERMANOVA,  $R^2 = 0.098$ ,  $p = 0.617$ ) (Fig.3c). Moreover, mechanical lysing time showed a statistical difference between samples (PERMANOVA,  $R^2 = 0.018$ ,  $p = 0.025$ ). Therefore, the effect size is bigger for PCR protocols and mechanical lysing time compared to extraction protocols. Of note, the two MP3 protocols ( $4 \times 5$  min bead beating) are situated distant from their respective groups (stratified by PCR protocol), and approaching the theoretical distribution (Fig.3c).

# ChatGPT – Journal articles

---

Please discuss the following results “ RESULTS “:

I chose the section “DNA extraction and PCR protocols impact 16S rRNA gene analysis” from VILLETTE et al.

DISCUSSION PROPOSED:

The results of this study suggest that the methods used for DNA extraction and mechanical lysing have a significant impact on the accuracy of high-throughput phylogenetic analysis of the 16S rRNA gene in low biomass samples. The study assessed three DNA extraction protocols and tested different mechanical lysing programs on microbial standards containing 8 fully characterized and equally abundant strains. The results show that overall, alpha diversity (a measure of the variety of different species within a sample) increased with the extend of mechanical lysing. Nested PCR (a variation of PCR) resulted in an increased alpha-diversity compared to standard PCR.

In addition, the study found that mechanical lysing was of great importance for the quality of extraction, as increasing mechanical lysing time and repetition improved the experimental outcome. More precisely extended mechanical lysing improved the extraction (relative abundance) of bacteria such as *B. subtilis* and *L. adolescentis*, over *E. coli* and *S. enterica*. Based on the distance to the theoretical composition, the study identified the optimal extraction protocol as the Miniprep kit with 4 × 5 min of bead beating for both PCR protocols tested. The results show that the PCR protocols also significantly impact the distance between experimental and theoretical composition, the Nested PCR protocol resulted in a lower Bray–Curtis distance than the Standard PCR protocol.

Principal coordinate analysis based on  $\beta$ -diversity showed two distinct clusters based on PCR protocols, but not based on extraction protocols. The study found that mechanical lysing time showed a statistical difference between samples. Overall, the results suggest that the effect size of PCR protocols and mechanical lysing time is bigger compared to extraction protocols. The study concludes that the optimal protocol for robust and reproducible microbiota analysis is the Miniprep kit with 4 × 5 min of bead beating and nested PCR.

This does not work well – ChatGPT cannot include references and it cannot compare to literature because it doesn't have access to all the primary litterature. However, you can feed a reference article to the ChatGPT, which will allow it to compare, however only based on what is written (no data analysis based on figures obviously).

# ChatGPT – Journal articles

---

Please discuss the following results “ RESULTS “:

I chose the section “DNA extraction and PCR protocols impact 16S rRNA gene analysis” from VILLETTE et al.

DISCUSSION PROPOSED:

The results of this study suggest that the methods used for DNA extraction and mechanical lysing have a significant impact on the accuracy of high-throughput phylogenetic analysis of the 16S rRNA gene in low biomass samples. The study assessed three DNA extraction protocols and tested different mechanical lysing programs on microbial standards containing 8 fully characterized and equally abundant strains. The results show that overall, alpha diversity (a measure of the variety of different species within a sample) increased with the extend of mechanical lysing. Nested PCR (a variation of PCR) resulted in an increased alpha-diversity compared to standard PCR.

In addition, the study found that mechanical lysing was of great importance for the quality of extraction, as increasing mechanical lysing time and repetition improved the experimental outcome. More precisely extended mechanical lysing improved the extraction (relative abundance) of bacteria such as *B. subtilis* and *L. adolescentis*, over *E. coli* and *S. enterica*. Based on the distance to the theoretical composition, the study identified the optimal extraction protocol as the Miniprep kit with 4 × 5 min of bead beating for both PCR protocols tested. The results show that the PCR protocols also significantly impact the distance between experimental and theoretical composition, the Nested PCR protocol resulted in a lower Bray–Curtis distance than the Standard PCR protocol.

Principal coordinate analysis based on  $\beta$ -diversity showed two distinct clusters based on PCR protocols, but not based on extraction protocols. The study found that mechanical lysing time showed a statistical difference between samples. Overall, the results suggest that the effect size of PCR protocols and mechanical lysing time is bigger compared to extraction protocols. The study concludes that the optimal protocol for robust and reproducible microbiota analysis is the Miniprep kit with 4 × 5 min of bead beating and nested PCR.

This does not work well – ChatGPT cannot include references and it cannot compare to literature because it doesn't have access to all the primary literature. However, you can feed a reference article to the ChatGPT, which will allow it to compare, however only based on what is written (no data analysis based on figures obviously).



## scientific reports

OPEN

### Refinement of 16S rRNA gene analysis for low biomass biospecimens

Remy Villette<sup>1,5</sup>, Gaëlle Autaa<sup>1,5</sup>, Sophie Hind<sup>2</sup>, Johanna B. Holm<sup>3</sup>, Alicia Moreno-Sabater<sup>1,2,4</sup> & Martin Larsen<sup>1,2,5</sup>

High-throughput phylogenetic 16S rRNA gene analysis has permitted to thoroughly delve into microbial community complexity and to understand host-microbiota interactions in health and disease. The analysis comprises sample collection and storage, genomic DNA extraction, 16S rRNA gene amplification, high-throughput amplicon sequencing and bioinformatic analysis. Low biomass microbiota samples (e.g. biopsies, tissue swabs and lavages) are receiving increasing attention, but optimal standardization for analysis of low biomass samples has yet to be developed. Here we tested the lower bacterial concentration required to perform 16S rRNA gene analysis using three different DNA extraction protocols, three different mechanical lysing series and two different PCR protocols. A mock microbiota community standard and low biomass samples ( $10^3$ ,  $10^7$ ,  $10^8$ ,  $10^9$  and  $10^{10}$  microbes) from two healthy donor stools were employed to assess optimal sample processing for 16S rRNA gene analysis using paired-end Illumina MiSeq technology. Three DNA extraction protocols tested in our study performed similar with regards to representing microbiota composition, but extraction yield was better for silica columns compared to bead absorption and chemical precipitation. Furthermore, increasing mechanical lysing time and repetition did ameliorate the representation of bacterial composition. The most influential factor enabling appropriate representation of microbiota composition remains sample biomass. Indeed, bacterial densities below  $10^6$  cells resulted in loss of sample identity based on cluster analysis for all tested protocols. Finally, we excluded DNA extraction bias using a genomic DNA standard, which revealed that a semi-nested PCR protocol represented microbiota composition better than classical PCR. Based on our results, starting material concentration is an important limiting factor, highlighting the need to adapt protocols for dealing with low biomass samples. Our study suggests that the use of prolonged mechanical lysing, silica membrane DNA isolation and a semi-nested PCR protocol improve the analysis of low biomass samples. Using the improved protocol we report a lower limit of  $10^6$  bacteria per sample for robust and reproducible microbiota analysis.

Lets try to ask ChatGPT to:

- propose titles ✓
- Improve abstract ✓
- Rewrite an abstract for non-experts (different age groups) ✓
- Write an introduction based on abstract (Need guidelines and cannot add references)
- Write a discussion based on a result section. (Cannot compare with previous literature)

# ChatGPT – Novel research projects

---

ChatGPT input:

Please provide a novel research project in the field of immune ageing

Could you explain why you think this project is novel.

ChatGPT result:

One potential research project in the field of immune aging could be to investigate the role of **epigenetics in the decline of immune function with age**. Epigenetics refers to changes in gene activity that do not involve changes to the underlying DNA sequence, and has been shown to play a role in aging and age-related diseases. The project could involve analyzing the epigenetic modifications in immune cells from young and old individuals, and determining how these changes affect the function of the immune system. Additionally, the project could involve investigating potential interventions, such as dietary or environmental factors, that may be able to reverse or slow down these epigenetic changes and improve immune function in older individuals.

This project is novel because it proposes to investigate the role of epigenetics in immune aging, which is an area of research that is still relatively understudied. **While the decline of immune function with age is well-established, the underlying mechanisms are not fully understood.** Epigenetics, which includes changes in gene activity that do not involve changes to the underlying DNA sequence, has been shown to play a role in aging and age-related diseases, but **its specific effects on the immune system have not been extensively studied.**

Additionally, the proposed project would investigate potential interventions, such as dietary or environmental factors,.....

# ChatGPT – Novel research projects

---

## ChatGPT input:

Could you add a specific cellular pathway which would be relevant to study in the context of this project.

## ChatGPT result:

One specific cellular pathway that would be relevant to study in the context of this project is the **DNA methylation pathway**. DNA methylation is an epigenetic modification that involves the addition of a methyl group to the DNA molecule, and is known to play a role in aging and age-related diseases. **This pathway is interesting to study in the context of immune aging because it has been shown to regulate the expression of genes involved in the immune response, such as those encoding cytokines and receptors.**

In this project, the researchers could analyze the levels of **DNA methylation in immune cells from young and old individuals**, and investigate whether there are any specific changes in the methylation patterns that are associated with age-related decline in immune function. They could also investigate whether these changes in **methylation patterns are specific to certain types of immune cells**, such as T cells or B cells, or whether they are more general changes that affect multiple immune cell types.

Additionally, the researchers could study the **effects of interventions such as dietary supplements or environmental factors on the DNA methylation pathway** in immune cells, to determine whether these interventions can reverse or slow down age-related changes in methylation patterns and improve immune function in older individuals..

# ChatGPT – Teaching - La cueillette by Charles Cros

---

Fait un commentaire composé du poème suivant en français:

C'était un vrai petit voyou,  
Elle venait on ne sait d'où,  
Moi, je l'aimais comme une bête.  
Oh ! la jeunesse, quelle fête.

Un baiser derrière son cou  
La fit rire et me rendit fou.  
Sainfoin, bouton d'or, pâquerette,  
Surveillaient notre tête à tête.

La clairière est comme un salon  
Tout doré ; les jaunes abeilles  
Vont aux fleurs qui leur sont pareilles ;

Moi seul, féroce et noir frelon,  
Qui baise ses lèvres vermeilles,  
Je fais tache en ce fouillis blond

ChatGPT result:

Le poème "**C'était un vrai petit voyou**" de Charles Cros est une description romantique et nostalgique de l'amour jeune. **Le narrateur se remémore une relation passée avec une jeune femme** qui était "un vrai petit voyou", et qu'il aimait "comme une bête". Les **images utilisées dans le poème évoquent un sentiment de jeunesse insouciance, avec le narrateur et son amoureuse** qui se délectent de la compagnie l'un de l'autre dans une prairie dorée entourée de fleurs sauvages. Le **narrateur utilise également une métaphore** en se comparant à un "frelon noir et féroce" parmi le fouillis blond des fleurs, mettant en évidence son amour passionné, intense et presque dangereux pour elle. **Le poème a un ton mélancolique et nostalgique**, il parle de la beauté de la jeunesse et de la nature éphémère de l'amour.

A really good proposition, which can at least start your though process and give an idea of what to include. The title is wrong, because it uses the first line as title and I didn't provide the title in the first line.

# ChatGPT – Teaching - Music

---

Write some jazz music

ChatGPT result:

I'm sorry, but as a language model, I am not able to generate music by myself, I can only provide you with text information and examples of music in written form. Jazz music is an art form that requires a high level of skill, creativity, and practice. It is composed and performed by musicians who have dedicated their lives to mastering their craft. It would be best to find a professional jazz musician or composer to create jazz music that you are looking for.

Creative limit

# ChatGPT – Teaching - Music

---

Write song lyrics for a love song

Creative capacity

ChatGPT result:

Verse 1:

I never knew what love was  
Till the day I met you  
You opened my eyes  
And my heart too

Chorus:

With you by my side  
I'm never alone  
Together we'll chase  
Our dreams to the unknown

Verse 2:

In your arms I feel alive  
With every kiss and every sigh  
I know that this is right  
You're my reason to fight

Chorus:

With you by my side  
I'm never alone  
Together we'll chase  
Our dreams to the unknown

Bridge:

I promise to love you  
With every beat of my heart  
I'll be there to hold you  
From now till forever starts

Chorus:

With you by my side  
I'm never alone  
Together we'll chase  
Our dreams to the unknown

Outro:

And I'll love you  
Till the end of time  
With you by my side  
Everything is just fine.



# ChatGPT – Teaching - Music

---

Write chord progression for a melody for the song you suggested.

## ChatGPT result:

Sure, here is a chord progression that could be used for the melody of the love song lyrics I provided earlier:

Verse:

Am - C - G - F

Chorus:

F - C - G - Am

Verse 2:

G - D - Am - F

Chorus:

F - C - G - Am

Bridge:

C - G - Am - F

Chorus:

F - C - G - Am

Outro:

Am - F - C - G

You can play around with the progression and add extra chords to make it more interesting or change it to match the melody you have in mind. These chords are just a suggestion, you can use different chords that you think will fit better with the melody and the overall feeling of the song.

# ChatGPT – Teaching - Music

---

Write the melody for the chorus.

## ChatGPT result:

Sure, here is a possible melody for the chorus of the love song lyrics I provided earlier:

F C G Am  
E D F G  
F C G Am  
E D F G

Please note that this is just one possible melody, and you can add your own touch to it or change it to match the style you are looking for. The melody should match the rhythm and the words of the song you have in mind.

Also, keep in mind that this is just a melody in written form, to play it you should have knowledge of music notation and be able to play an instrument.

# ChatGPT – Teaching - Music

---

Which key did you write in?

## ChatGPT result:

The melody and chord progression I provided were written in the key of F.

The chords that I provided: F, C, G, Am, D, are all diatonic chords in the key of F major. This means that all the chords are built from the notes of the F major scale.

You can change the key to better match your vocal range or the instruments you are going to play it with.

Keep in mind that this is just a suggestion and you can experiment with different keys and chord progressions to find the one that best suits your song.

# ChatGPT – Teaching - Music

---

You wrote the melody as a code. Can you propose a website where I can play it?

ChatGPT result:

I apologize, I misunderstood your question earlier. I provided the melody in a written form in sheet music notation, which uses symbols such as notes, clefs, and bars, to represent music.

There are many websites and software where you can input sheet music notation and play the melody. Some of the most popular ones are :

**MuseScore:** A free and open-source software for composing, playing, and printing sheet music.

**Virtual Piano:** A web-based virtual piano that allows you to play and hear the melody using your computer keyboard.

**Flat.io:** A web-based tool that allows you to compose, play, and share sheet music.

**Noteflight:** A web-based tool that allows you to compose, play, and print sheet music.

These are just a few examples, and you can find many more by searching for "online sheet music player" or "sheet music notation software" in your preferred search engine.

Once again, I apologize for any confusion caused.

It “thinks” that when I asked for the key that I was talking about this question. It therefore apologizes for the misunderstanding. In other words all questions are considered in the context of the chat.

# ChatGPT – Teaching - Music

Write the melody for the chorus.

## ChatGPT result:

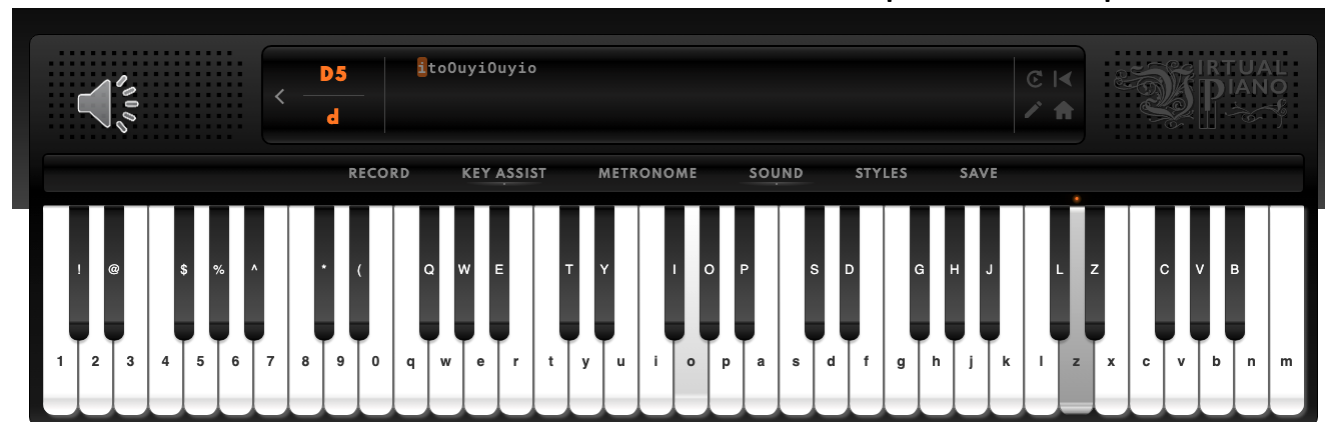
Sure, here is a possible melody for the chorus of the love song lyrics I provided earlier:

F C G Am  
E D F G  
F C G Am  
E D F G

Please note that this is just one possible melody, and you can add your own touch to it or change it to match the style you are looking for. The melody should match the rhythm and the words of the song you have in mind.

Also, keep in mind that this is just a melody in written form, to play it you should have knowledge of music notation and be able to play an instrument.

<https://virtualpiano.net/>



# ChatGPT – Teaching - History

---

ChatGPT input:

Explique debut de la premiere guerre mondiale

ChatGPT result:

The outbreak of the First World War can be traced back to the **assassination of Archduke Franz Ferdinand of Austria on June 28, 1914**. The assassination, which was carried out by a member of a Bosnian



# ChatGPT – Teaching - History

---

ChatGPT input:

Explique debut de la premiere guerre mondiale en francais

fait plus court

encore plus court

ChatGPT result:

Le début de la Première Guerre mondiale peut être remonté à **l'assassinat de l'archiduc François-Ferdinand d'Autriche le 28 juin 1914**. L'assassinat, perpétré par un membre d'un groupe nationaliste bosniaque, a conduit à une série d'événements qui ont finalement déclenché la guerre. Les tensions entre les grandes puissances européennes, qui étaient divisées en deux blocs rivaux (les Triple-Entente et les Puissances centrales), ont rapidement augmenté après l'assassinat. **Les alliances militaires et les traités de défense précédemment conclus ont conduit à une escalation rapide des hostilités, avec de nombreux pays entrant rapidement dans la guerre.** La Première Guerre mondiale a finalement impliqué des pays de tous les continents, et a été l'un des conflits les plus meurtriers de l'histoire, causant la mort de millions de personnes.

La Première Guerre mondiale a débuté suite à l'assassinat de l'archiduc François-Ferdinand d'Autriche en 1914. Les tensions entre les grandes puissances européennes, divisées en deux blocs rivaux, ont rapidement augmenté après cet événement. **Les alliances et les traités de défense ont conduit à une escalation rapide des hostilités**, impliquant des pays de tous les continents et causant des millions de morts.

La Première Guerre mondiale a débuté après l'assassinat de l'archiduc François-Ferdinand en 1914, les tensions entre les grandes puissances européennes ayant conduit à une escalation rapide et globale des hostilités. Elle a causé des millions de morts.

# ChatGPT – Teaching - Economy

---

Can you predict Tesla's stock price in the coming week?

ChatGPT result:

**I am a language model AI, I don't have the ability to predict future stock prices** as it's influenced by many variables and factors such as market conditions, economic indicators, company performance, and global events, among others. The stock market is highly unpredictable and making predictions about future stock prices is difficult, even for experts. It's important to conduct thorough research and consult with financial experts before making any investment decisions. It's also worth noting that past performance of a stock doesn't guarantee future results.

At least not leading people into economic disaster

# ChatGPT – Teaching - Biology

---

Please describe the mechanisms of bacterial motility

## ChatGPT result:

**Bacterial motility** refers to the ability of bacteria to move and navigate their environment. **There are several mechanisms** by which bacteria can achieve motility, including:

1. **Flagella-based motility:** Many bacteria are equipped with one or more whip-like appendages called flagella. These flagella are used for propulsion, allowing the bacteria to swim through liquids. The flagella rotate in a coordinated fashion, propelling the bacteria in a specific direction.
2. **Pili-based motility:** Some bacteria use hair-like appendages called pili for motility. Pili are used in a process called twitching motility, where the bacteria extend and retract their pili in a coordinated fashion, allowing them to pull themselves along surfaces.
3. **Gliding motility:** Some bacteria are able to move across surfaces without the use of flagella or pili. This is called gliding motility, and the mechanism by which this occurs is not fully understood. Some researchers believe that gliding motility is driven by the secretion of slime or other extracellular materials, while others propose that it is driven by the interaction of the bacterial cell surface with the surface it is gliding on.
4. **Chemotaxis:** Bacteria are also able to sense and respond to chemical gradients in their environment, a process known as chemotaxis. This allows them to move towards or away from specific chemicals. They do this by using transmembrane receptors that can detect a chemical and use the energy of the ATP hydrolysis to move their flagella or pili in a coordinated manner.
5. **Phototaxis:** Phototaxis is the ability of bacteria to move in response to light. Bacteria use light receptors on their membrane to sense the direction and intensity of light. They then adjust their flagella or pili accordingly to move towards or away from the light source.

All of these mechanisms allow bacteria to navigate their environment, find nutrients, and evade predators or other potential threats.

This is a good and fairly comprehensive answer! One can ask for further details or a reduced answer.

# ChatGPT – Teaching - Biology

Please explain the difference between mitosis and meiosis:

## ChatGPT result:

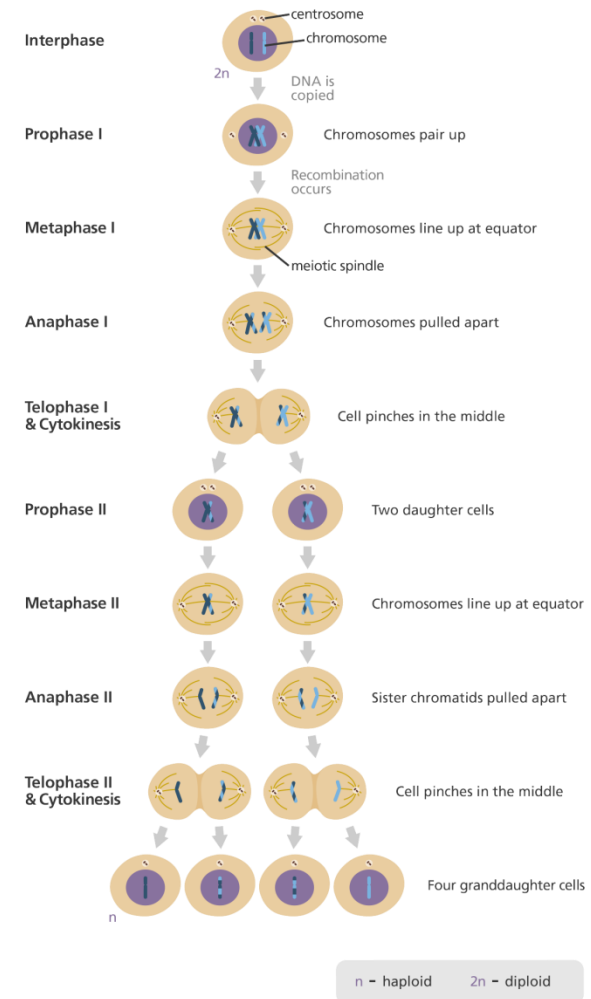
**Mitosis and meiosis are both cellular processes that involve the division of a cell's nucleus and the separation of chromosomes.** However, they serve different purposes and have distinct characteristics.

**Mitosis** is a type of cell division that results in the formation of two identical daughter cells, each with the same number of chromosomes as the parent cell. Mitosis is responsible for the growth and repair of tissues in multicellular organisms. It takes place in somatic cells, which are cells that make up the body of an organism and are **not involved in sexual reproduction**.

**Meiosis**, on the other hand, is a specialized form of cell division that results in the formation of **four genetically diverse daughter cells**, each with half the number of chromosomes as the parent cell. Meiosis is responsible for the production of sperm and egg cells in sexually reproducing organisms. These daughter cells are called gametes, which fuse during fertilization to form a zygote with a full set of chromosomes.

**In summary**, mitosis is the process by which cells divide to form two identical daughter cells for growth and repair, while meiosis is the process by which cells divide to form genetically diverse daughter cells for sexual reproduction.

This is a good and fairly comprehensive answer! One can ask for further details or a reduced answer.



# ChatGPT – Teaching - Math

---

ChatGPT input:

Create the derivative of  $y=2x^2+4x-5$  and find solutions for  $y=0$ :

Here it stopped midway in the quadratic formula. Serious problems with math calculations, although the theory is available and it can explain the theory in simple terms.

Of note, ChatGPT often stops midway and you can ask it to finalize it's previous response. I think this is primarily due to server incapacity.

ChatGPT result:

To find the derivative of  $y = 2x^2 + 4x - 5$ , we can use the power rule of calculus which states that:

$$d/dx (x^n) = nx^{(n-1)}$$

so the derivative of  $y = 2x^2 + 4x - 5$  with respect to  $x$  is:

$$\begin{aligned} dy/dx &= d/dx (2x^2) + d/dx (4x) + d/dx (-5) \\ &= 2 \cdot 2x^{(2-1)} + 1 \cdot 4x^{(1-1)} + 0 \cdot (-5) \\ &= 4x + 4 \end{aligned}$$

So, the derivative of  $y = 2x^2 + 4x - 5$  is  $4x + 4$

To find the solutions for  $y = 0$ , we can set the original equation equal to 0 and solve for  $x$ . So, we have:

$$2x^2 + 4x - 5 = 0$$

To solve this equation we can either factor it or use the quadratic formula.

Factoring it:

$$\begin{aligned} (2x+1)(x-5) &= 0 \\ x &= -1/2 \text{ or } x = 5 \end{aligned}$$

Using the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$x = \frac{-4 \pm \sqrt{4^2 - 4 \cdot 2 \cdot (-5)}}{2 \cdot 2}$$

# ChatGPT – Teaching - Math

---

ChatGPT input:

Find the minimum value of y:

Here it makes a mistake for the second derivative, but we can ask ChatGPT to correct itself.

ChatGPT result:

To find the minimum value of y, we first need to find the critical points of the function, that is, the points where the derivative of y is equal to 0 or is undefined.

In this case, the derivative of  $y = 2x^2 + 4x - 5$  is  $4x + 4$ .

So, the critical points of the function are:

$$4x + 4 = 0$$

$$x = -1$$

This means that  $x = -1$  is a critical point of the function.

Now, we will use the second derivative test, it consists of computing the second derivative of the function and then evaluating it at the critical point. if the second derivative is positive at the critical point, then the function has a local minimum at that point.

The second derivative of y is:

$$d^2y/dx^2 = d/dx (4x + 4) = d/dx (4) = 0$$

Since the second derivative is equal to zero, we can't confirm if the function has a local minimum or not, we can only conclude that there is a point of inflection at  $x = -1$

So, there is no minimum value for  $y = 2x^2 + 4x - 5$ .

# ChatGPT – Teaching - Math

---

ChatGPT input:

Your second derivative is not equal to 0, but equal to 4. Correct your error please:

It makes another mistake, because it doesn't know that  $(-1)^2$  is 1 and not -1. We correct it again.

ChatGPT result:

You are correct, my apologies for the confusion. I apologize for the confusion, the second derivative of  $y = 2x^2 + 4x - 5$  is indeed 4. Since the second derivative is positive, it means that the function has a local minimum at the critical point.

The second derivative of y is:

$$d^2y/dx^2 = d/dx (4x + 4) = d/dx (4) = 4$$

Since the second derivative is positive, it means that the function has a local minimum at the critical point  $x=-1$

To find the minimum value of y, we can substitute  $x=-1$  into the original function  $y = 2x^2 + 4x - 5$

$$y = 2(-1)^2 + 4(-1) - 5$$

$$y = -2 + (-4) - 5$$

$$y = -11$$

So, the minimum value of y is **-11**.



# ChatGPT – Teaching - Math

---

ChatGPT input:

No the minimum value is -7:

It gets the right result finally. However, it has been remarked by users that ChatGPT will always agree with a person if he/she corrects the ChatGPT, even if this person is wrong. At this stage it is a good tool, but you basically need to know the answer to the question better than the ChatGPT, which is obviously not the purpose.

I have tested whether it always agree with the user. I asked:

Are men more intelligent than women?

The ChatGPT responds with a no and explains why.

If I insist and say that “men are more intelligent than women” it stands firm on its point that there are no scientific evidence to backup my claim.

ChatGPT result:

I apologize for the confusion, you are correct. I apologize for the mistake on my previous answer. To find the minimum value of  $y$ , we can substitute  $x=-1$  into the original function  $y = 2x^2 + 4x - 5$

$$y = 2(-1)^2 + 4(-1) - 5$$

$$y = 2 + (-4) - 5$$

$$y = -7$$

so the minimum value of  $y$  is -7.

# ChatGPT – Teaching – Math – Matlab model

---

ChatGPT input:

Can you write a matlab code representing the dynamics of an ecological system of foxes and rabbits:

ChatGPT result:

I can certainly provide an example of a Matlab code that models the dynamics of an ecological system of foxes and rabbits, but it is important to note that this model will be a simplification of a real-world ecosystem and may not accurately represent the behavior of foxes and rabbits in nature. Here is an example of a **Lotka-Volterra model of the fox-rabbit system**:

% Define parameters

alpha = 0.1; % rabbit birth rate

beta = 0.02; % rabbit death rate due to predation

gamma = 0.3; % fox death rate

delta = 0.01; % fox birth rate

% Define initial populations

x0 = [100; 20]; % initial rabbit and fox populations

% Define time span for simulation

tspan = [0, 50]; % simulate for 50 time units

% Define the system of differential equations

f = @(t,x) [alpha\*x(1) - beta\*x(1)\*x(2); delta\*x(2) - gamma\*x(1)\*x(2)];

% Use Matlab's ODE solver to simulate the system

[t,x] = ode45(f, tspan, x0);

% Plot the results

plot(t,x(:,1),'r',t,x(:,2),'g')

legend('Rabbits','Foxes')

xlabel('Time')

ylabel('Population')

This code defines the parameters of the model, including rabbit birth and death rates and fox birth and death rates. It then defines the initial populations of rabbits and foxes and the time span for the simulation. The system of differential equations representing the dynamics of the fox-rabbit system is defined using a function handle, and Matlab's ODE45 solver is used to simulate the system. Finally, the results are plotted with time on the x-axis and population on the y-axis.

This is just an example, you need to adjust the parameters to a realistic value and also this is a simple model, in a real world scenario the ecosystem is more complex and there might be other factors involved such as diseases, availability of food, hunting, etc.

# ChatGPT – Teaching – Statistics – T test

---

ChatGPT input:

I have two independent variables X1 and X2. Please make an R script, which tests all conditions required to make a t test (compare means) and add graphics representing the data. Density plots for X1 and X2 and a violin plot for comparison. Include p value in violin plot.

ChatGPT result:

```
# load required libraries
library(ggplot2)
library(ggpubr)
```

```
# generate example data
set.seed(123)
X1 <- rnorm(100, mean = 5, sd = 1)
X2 <- rnorm(100, mean = 7, sd = 2)
```

```
# test for normality of X1 and X2
shapiro.test(X1)
shapiro.test(X2)
```

```
# test for equal variances
var.test(X1, X2)
```

```
# perform t-test
t.test(X1, X2, var.equal = TRUE)
```

```
# create density plots for X1 and X2
ggdensity(X1, xlab = "X1", fill = "red") + ggdensity(X2, xlab = "X2", fill = "blue")
```

```
# create violin plot for comparison with p-value displayed
data <- data.frame(X1 = X1, X2 = X2)
p <- ggviolin(data, x = "X1", y = "X2", add = "boxplot",
              boxfill = "white", violinfill = "blue",
              ylab = "X", xlab = "Group")
p + stat_compare_means(comparisons = list(c("X1", "X2")),
                       label = "p-value")
```

# ChatGPT – Teaching – Statistics – T test

ChatGPT input:

I have two independent variables X1 and X2. Please make an R script, which tests all conditions required to make a t test (compare means) and add graphics representing the data. Density plots for X1 and X2 and a violin plot for comparison. Include p value in violin plot.

ChatGPT result:

The code is wrong, because ggplot expects the data format to be long. See major change in **bold**.

X1 X2		pivot_longer →	Group value	
X1	X2		Group	value
7.20	7.24		X1	7.20
6.31	7.49		X1	6.31
4.73	9.46		X1	4.73
5.54	5.97		X1	5.54
4.59	5.01		X1	4.59
4.52	10.35		X1	4.52
4.21	6.12		X1	4.21
4.41	5.55		X1	4.41
6.65	4.53		X1	6.65
4.95	4.43		X1	4.95
			X2	7.24
			X2	7.49
			X2	9.46
			X2	5.97
			X2	5.01
			X2	10.35
			X2	6.12
			X2	5.55
			X2	4.53
			X2	4.43

```
# perform t-test
t.test(X1, X2, var.equal = TRUE)
```

```
# create density plots for X1 and X2
ggdensity(X1, xlab = "X1", fill = "red") + ggdensity(X2, xlab = "X2", fill = "blue")
```

```
# create violin plot for comparison with p-value displayed
data <- data.frame(X1 = X1, X2 = X2)
data_long <- pivot_longer(data, cols=c("X1","X2"), names_to="Group")
p <- ggplot(data=data_long, aes(x=Group, y=value))+
  geom_violin(aes(fill=Group),alpha=0.4)+
  geom_boxplot(width=.1)+
  geom_jitter(aes(color=Group), alpha=0.8, size=3)+
  labs(xlab="Group",ylab="X")+
  theme_classic2()+
  stat_compare_means(comparisons = list(c("X1", "X2")), label = "p-value")
```

# ChatGPT – Teaching – Statistics – T test

ChatGPT input:

I have two independent variables X1 and X2. Please make an R script, which tests all conditions required to make a t test (compare means) and add graphics representing the data. Density plots for X1 and X2 and a violin plot for comparison. Include p value in violin plot.

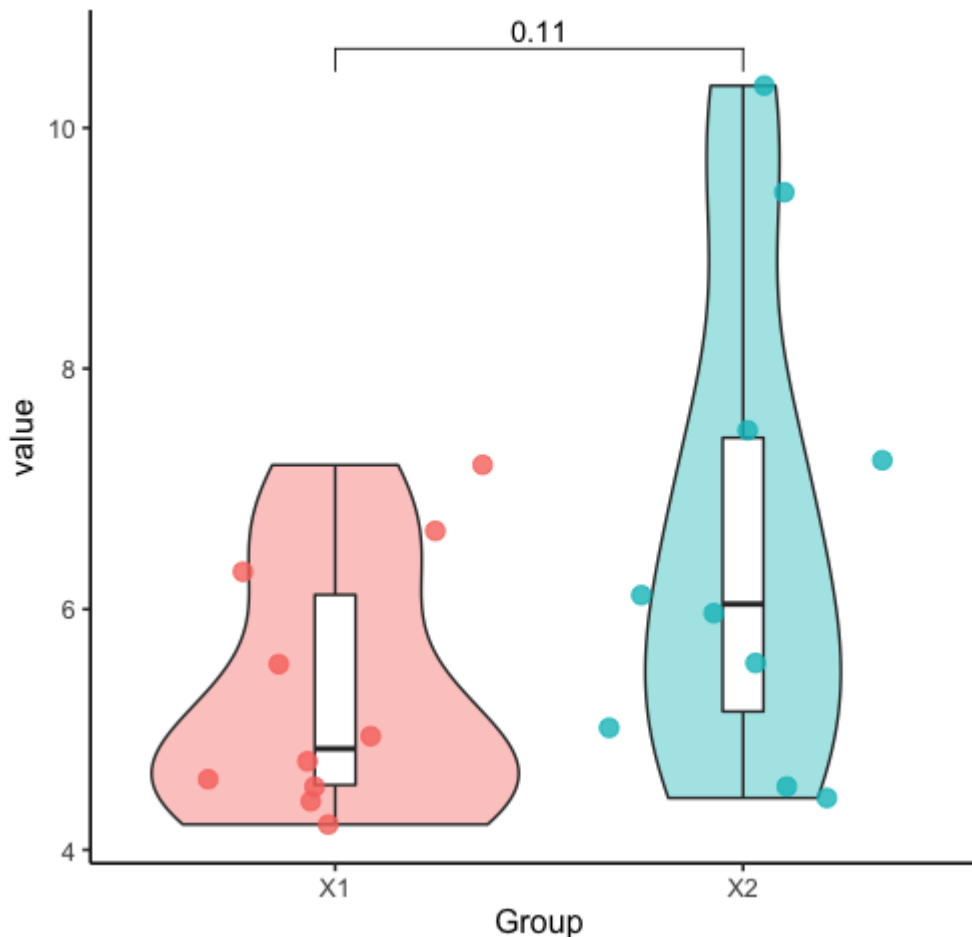
ChatGPT result

```
# load required libra
library(ggplot2)
library(ggpubr)
library(tidyr)
```

```
# generate example
set.seed(123)
X1 <- rnorm(100, m
X2 <- rnorm(100, m
```

```
# test for normality
shapiro.test(X1)
shapiro.test(X2)
```

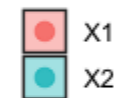
```
# test for equal vari
var.test(X1, X2)
```



```
ggdensity(X2, xlab =
```

```
alue displayed
```

Group



```
cols=c("X1", "X2"),
```

```
y=value))+
```

```
, size=3))+
```

```
(c("X1", "X2")), label
```

# ChatGPT – Teaching – Statistics – T test

## ChatGPT input:

Please transform script into an R shiny app, which allows you to select mean and standard deviation for X1 and X2 using sliders and adds the option of paired versus unpaired analysis. Please also use the var.test result to determine if var.equal should be set to true or false.

## ChatGPT result:

```
# load required libraries
library(shiny)
library(ggplot2)
library(ggpubr)
set.seed(123)
```

```
# define UI
ui <- fluidPage(
  titlePanel("T-Test Comparison"),

  sidebarLayout(
    sidebarPanel(
      sliderInput("meanX1", "Mean of X1", min = 0, max = 10,
value = 5),
      sliderInput("sdX1", "Standard Deviation of X1", min = 0,
max = 5, value = 1),
      sliderInput("meanX2", "Mean of X2", min = 0, max = 10,
value = 7),
      sliderInput("sdX2", "Standard Deviation of X2", min = 0,
max = 5, value = 2),
      radioButtons("paired", "Paired or Unpaired?", c("Paired"
= "paired", "Unpaired" = "unpaired"), selected = "unpaired"),
    ),
```

```
mainPanel(
  plotOutput("densityPlot"),
  plotOutput("violinPlot"),
)
```

```
)
)
)

# define server logic
server <- function(input, output) {
```

```
  # generate data
  data <- reactive({
    if (input$paired=="paired"){
      n1=input$n1
      n2=input$n2
    } else {
      n1=input$n1
      n2=input$n2
    }
    X1 <- rnorm(n1, mean = input$meanX1, sd = input$sdX1)
    X2 <- rnorm(n2, mean = input$meanX2, sd = input$sdX2)
    X1 <- data.frame(Group="X1", value=X1)
    X2 <- data.frame(Group="X2", value=X2)
    return(as.data.frame(rbind(X1,X2)))
  })
```

```
  # create density plot
  output$densityPlot1 <- renderPlot({
    data <- data()
    ggdensity(data[data$Group=="X1","value"], xlab = "X1",
fill = "red")
  }, width = 300, height = 300)
```

```
data <- data()
ggdensity(data[data$Group=="X2","value"], xlab = "X2",
fill = "blue")
}, width = 300, height = 300)
```

```
# create violin plot
output$violinPlot <- renderPlot({
  var_test <- var.test(data()$X1, data()$X2)
  p <- ggviolin(data(), x = "X1", y = "X2", add = "boxplot",
    boxfill = "white", violinfill = "blue",
    ylab = "X", xlab = "Group")
  p + stat_compare_means(comparisons = list(c("X1",
"X2")), label = "p-value")
})
```

```
# perform t-test
output$tTestResult <- renderPrint({
  if(input$paired == "paired"){
    t.test(data()$X1, data()$X2, paired = TRUE)
  }else{
    t.test(data()$X1, data()$X2, var.equal =
var_test$p.value > 0.05)
  }
})
```

```
# run the app
shinyApp(ui = ui, server = server)
```

# ChatGPT – Teaching – Statistics – T test

ChatGPT input:

Please transform script into an R shiny app, which allows you to select mean and standard deviation for X1 and X2 using sliders and adds the option of paired versus unpaired analysis. Please also use the `var.test` result to determine if `var.equal` should be set to true or false.

( The Shiny App code contains errors. E.g.

- `var_test` is not available when called in “tTestResult”.
- `violinPlot` is based on a wrong data structure
- It suddenly forgot to include the normality tests.

```
# define UI
ui <- fluidPage(
  titlePanel("T-Test Comparison"),

  sidebarLayout(
    sidebarPanel(
      sliderInput("meanX1", "Mean of X1", min = 0, max = 10,
value = 5),
      sliderInput("sdX1", "Standard Deviation of X1", min = 0,
max = 5, value = 1),
      sliderInput("meanX2", "Mean of X2", min = 0, max = 10,
value = 7),
      sliderInput("sdX2", "Standard Deviation of X2", min = 0,
max = 5, value = 2),
      radioButtons("paired", "Paired or Unpaired?", c("Paired"
= "paired", "Unpaired" = "unpaired"), selected = "unpaired"),
    ),

    # generate data
    data <- reactive({
      if (input$paired=="paired"){
        n1=input$n1
        n2=input$n2
      } else {
        n1=input$n1
        n2=input$n2
      }
      X1 <- rnorm(n1, mean = input$meanX1, sd = input$sdX1)
      X2 <- rnorm(n2, mean = input$meanX2, sd = input$sdX2)
      X1 <- data.frame(Group="X1", value=X1)
      X2 <- data.frame(Group="X2", value=X2)
      return(as.data.frame(rbind(X1,X2)))
    })

    # create density plot
    output$densityPlot1 <- renderPlot({
      data <- data()
      ggdensity(data[data$Group=="X1","value"], xlab = "X1",# run the app
fill = "red")
      }, width = 300, height = 300)

    # create violin plot
    output$violinPlot <- renderPlot({
      var_test <- var.test(data()$X1, data()$X2)
      p <- ggviolin(data(), x = "X1", y = "X2", add = "boxplot",
        boxfill = "white", violinfill = "blue",
        ylab = "X", xlab = "Group")
      p + stat_compare_means(comparisons = list(c("X1",
"X2")), label = "p-value")
    })

    # perform t-test
    output$tTestResult <- renderPrint({
      if(input$paired == "paired"){
        t.test(data()$X1, data()$X2, paired = TRUE)
      }else{
        t.test(data()$X1, data()$X2, var.equal =
var_test$p.value > 0.05)
      }
    })
  ),
  mainPanel(
    plotOutput("densityPlot"),
    plotOutput("violinPlot"),
  )
)
```



# ChatGPT – Teaching – Statistics – T test

---

A fairly rapid manual correction (15min) allowed to make a functional application.

Re-added shapiro test

Corrected violin plot to fit long data format

Corrected T test to fit long data format

I added tabulations for normality, homoscedasticity and T test.

I changed the sizes of the images.

# ChatGPT – Teaching – Statistics – T test

A fairly rapid manual correction (15min) allowed to make a functional application.

```
# load required libraries
library(shiny)
library(ggplot2)
library(ggpubr)
library(tidyr)
set.seed(123)

# define UI
ui <- fluidPage(
  titlePanel("T-Test Comparison"),

  sidebarLayout(
    sidebarPanel(
      conditionalPanel(condition = 'input.paired=="paired"',
        sliderInput("n", "Sample Size of X1 and X2", min = 0, max = 50, value = 10),
      ),
      conditionalPanel("input.paired=="unpaired",
        sliderInput("n1", "Sample Size of X1", min = 0, max = 50, value = 10),
      ),
      sliderInput("meanX1", "Mean of X1", min = 0, max = 10, value = 5),
      sliderInput("sdX1", "Standard Deviation of X1", min = 0, max = 5, value = 1),
      conditionalPanel("input.paired=="unpaired",
        sliderInput("n2", "Sample Size of X2", min = 0, max = 50, value = 10),
      ),
      sliderInput("meanX2", "Mean of X2", min = 0, max = 10, value = 7),
      sliderInput("sdX2", "Standard Deviation of X2", min = 0, max = 5, value = 1),
      radioButtons("paired", "Paired or Unpaired?", c("Paired" = "paired",
        "Unpaired" = "unpaired"), selected = "unpaired"),
    ),

    mainPanel(
      tabsetPanel(
        id="tabs",
        tabPanel("Normality",
          verbatimTextOutput("shapiro1"),
          plotOutput("densityPlot1"),
          verbatimTextOutput("shapiro2"),
          plotOutput("densityPlot2")
        ),
        tabPanel("Homoscedasticity",
          verbatimTextOutput("vartest")
        ),
        tabPanel("T test",

          verbatimTextOutput("tTestResult"),
          plotOutput("violinPlot")

        )
      )
    )
  )

# define server logic
server <- function(input, output) {

  # generate data
  data <- reactive({
    # print(input$paired)
    if (input$paired=="paired"){
      n1=input$n
      n2=n1
    } else {
      n1=input$n1
      n2=input$n2
    }
    X1 <- mnorm(n1, mean = input$meanX1, sd = input$sdX1)
    # print(X1)
    X2 <- mnorm(n2, mean = input$meanX2, sd = input$sdX2)
    X1 <- data.frame(Group="X1", value=X1)
    X2 <- data.frame(Group="X2", value=X2)
    data <- rbind(X1,X2)
    return(as.data.frame(data))
  })

  # Test for normality
  output$shapiro1 <- renderPrint({
    data <- data()
    if (!is.null(data)){
      shapiro.test(data[data$Group=="X1", "value"])
    }
  })
  output$shapiro2 <- renderPrint({
    data <- data()
    if (!is.null(data)){
      shapiro.test(data[data$Group=="X2", "value"])
    }
  })

  # create density plot
  output$densityPlot1 <- renderPlot({

    data <- data()
    ggdensity(data[data$Group=="X1", "value"], xlab = "X1", fill = "red")
  }, width = 300, height = 300)
  output$densityPlot2 <- renderPlot({
    data <- data()
    ggdensity(data[data$Group=="X2", "value"], xlab = "X2", fill = "blue")
  }, width = 300, height = 300)

  # test for equal variances
  output$varTest <- renderPrint({
    data <- data()
    var.test(data$value~data$Group)
  })

  # create violin plot
  output$violinPlot <- renderPlot({
    data <- data()
    p <- ggplot(data=data, aes(x=Group, y=value))+
      geom_violin(aes(fill=Group), alpha=0.4)+
      geom_boxplot(width=.1)+
      geom_jitter(aes(color=Group), alpha=0.8, size=3)+
      labs(xlab="Group", ylab="X")+
      theme_classic2()+
      stat_compare_means(comparisons = list(c("X1", "X2")), label = "p-value",
        method = "t.test", paired=(input$paired=="paired"))
    return(p)
  })

  # perform t-test
  output$tTestResult <- renderPrint({
    data <- data()
    var_test <- var.test(data$value~data$Group)
    vareq <- var_test$p.value > 0.05
    if(input$paired == "paired"){
      t.test(data$value~data$Group, paired = TRUE)
    }else{
      t.test(data$value~data$Group, var.equal=vareq)
    }
  })

  # run the app
  shinyApp(ui = ui, server = server)
}
```

# ChatGPT – Teaching – Statistics – T test

A fairly rapid manual correction (15min) allowed to make a functional application.

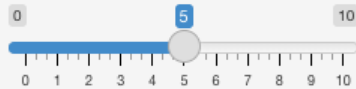
<http://127.0.0.1:5226> Open in Browser Publish

## T-Test Comparison

### Sample Size of X1



### Mean of X1



### Standard Deviation of X1



### Sample Size of X2



### Mean of X2



### Standard Deviation of X2



### Paired or Unpaired?

- ☐ Paired  
☒ Unpaired

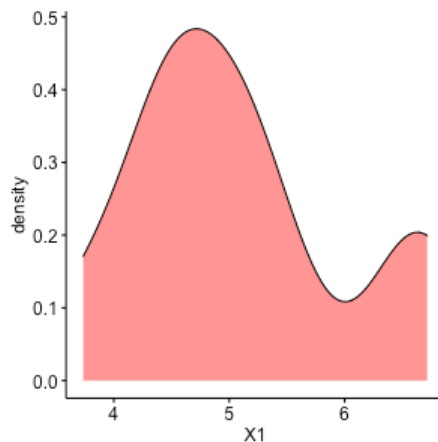
### Normality

### Homoscedasticity

### T test

#### Shapiro-Wilk normality test

```
data: data[data$Group == "X1", "value"]  
W = 0.92372, p-value = 0.389
```



#### Shapiro-Wilk normality test

```
data: data[data$Group == "X2", "value"]  
W = 0.95281, p-value = 0.7018
```

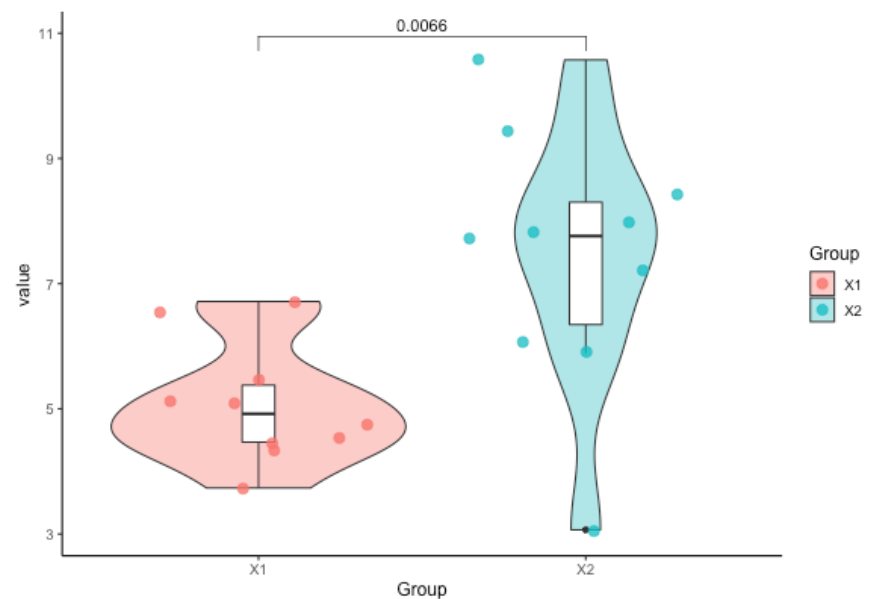
### Normality

### Homoscedasticity

### T test

#### Welch Two Sample t-test

```
data: data$value by data$Group  
t = -3.2424, df = 12.637, p-value = 0.006637  
alternative hypothesis: true difference in means between group X1 and group X2 is not equal to 0  
95 percent confidence interval:  
-3.9080601 -0.7771765  
sample estimates:  
mean in group X1 mean in group X2  
5.074626 7.417244
```



# ChatGPT – Teaching – Statistics - Homework

---

ChatGPT input:

Question 1. La vitesse de croissance de l'espèce *Sargassum muticum* suit une loi de distribution normale  $N(\mu=35 ; \sigma = 8,3)$  en  $\text{mm.j}^{-1}$ .

1) Calculez la probabilité qu'une algue de cette espèce présente une vitesse de croissance supérieure à  $43 \text{ mm.j}^{-1}$ .

2) Calculez la probabilité qu'un individu de cette espèce ait une vitesse de croissance inférieure à  $18 \text{ mm.j}^{-1}$  :

First two questions from a university level statistics exam. It cannot do the math, but proposes and explains the strategy for solving the problem.

ChatGPT result:

To calculate the probability that the growth rate of *Sargassum muticum* is greater than  $43 \text{ mm.j}^{-1}$ , we need to use the cumulative distribution function (CDF) of the normal distribution.

The CDF can be calculated using the following formula:

$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

To calculate the probability that growth rate is greater than  $43 \text{ mm.j}^{-1}$ , we will use the following formula :

$$P(X > 43) = 1 - P(X \leq 43) = 1 - F(43)$$

To calculate the probability that the growth rate of *Sargassum muticum* is less than  $18 \text{ mm.j}^{-1}$ , we can use the CDF as well.

The formula for this probability will be :

$$P(X < 18) = F(18)$$

In both cases, you need to use a calculator or software like R to calculate the CDF using the given mean and standard deviation.

It's important to note that these calculations are based on the assumption that the growth rate of *Sargassum muticum* follows a normal distribution with the given mean and standard deviation. If this assumption is not true in reality, these calculations may not be accurate.

# ChatGPT – Teaching – Statistics - Homework

---

ChatGPT input:

Please provide R code for the last question:

However, if you request R code to do the math it proposes working code.

**Code run in R studio:**

```
print(paste0("Probability that growth rate > 43: ", round(p1, 4)))  
[1] "Probability that growth rate > 43: 0.1676"  
> print(paste0("Probability that growth rate < 18: ", round(p2, 4)))  
[1] "Probability that growth rate < 18: 0.0203"
```

ChatGPT result:

Sure, here is the R code to calculate the probabilities in question 1:

```
# Load the library  
library(stats)
```

```
# Parameters  
mu <- 35  
sigma <- 8.3
```

```
# Probability that growth rate > 43  
p1 <- 1 - pnorm(43, mean = mu, sd = sigma)
```

```
# Probability that growth rate < 18  
p2 <- pnorm(18, mean = mu, sd = sigma)
```

```
# Print the results  
print(paste0("Probability that growth rate > 43: ", round(p1, 4)))  
print(paste0("Probability that growth rate < 18: ", round(p2, 4)))
```

In this code, it loads the stats library which contains the pnorm function to calculate the CDF of a normal distribution. Then it assigns the values of mu and sigma to the variables mu and sigma.

It uses the pnorm function with mean and sd as the mu and sigma values and the value of x as 43, 18 for the first and second question respectively. After that it calculates the probability using the formula  $p1 = 1 - \text{pnorm}(43, \text{mean} = \mu, \text{sd} = \sigma)$  and  $p2 = \text{pnorm}(18, \text{mean} = \mu, \text{sd} = \sigma)$ .

Finally, it prints the results rounded to 4 decimal places using the paste0 function.

Please note that the result of these calculations are based on the assumption that the growth rate of Sargassum muticum follows a normal distribution with the given mean and standard deviation. If this assumption is not true in reality, these calculations may not be accurate.

# ChatGPT – Teaching - Informatics

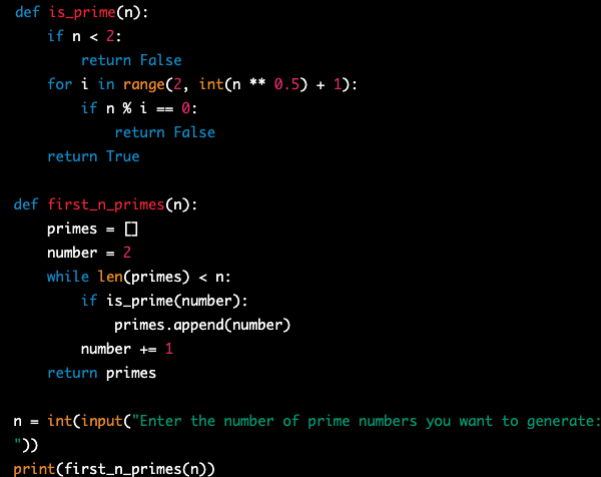
## Write a python program that will identify the first n prime numbers.

Here is an example of a Python program that will identify the first n prime numbers:

```
def is_prime(n):
    if n < 2:
        return False
    for i in range(2, int(n ** 0.5) + 1):
        if n % i == 0:
            return False
    return True

def first_n_primes(n):
    primes = []
    number = 2
    while len(primes) < n:
        if is_prime(number):
            primes.append(number)
        number += 1
    return primes

n = int(input("Enter the number of prime numbers you want to generate: "))
print(first_n_primes(n))
```

A screenshot of a code editor with a dark background and light-colored text. The code is a Python program to find the first n prime numbers. It includes a function 'is\_prime' that checks if a number is prime by testing divisibility from 2 to its square root. Another function 'first\_n\_primes' uses a while loop to find the first n primes. The main part of the code prompts the user for the number of primes to generate and prints the result. A 'Copy code' button is visible in the top right corner of the editor.

```
def is_prime(n):
    if n < 2:
        return False
    for i in range(2, int(n ** 0.5) + 1):
        if n % i == 0:
            return False
    return True

def first_n_primes(n):
    primes = []
    number = 2
    while len(primes) < n:
        if is_prime(number):
            primes.append(number)
        number += 1
    return primes

n = int(input("Enter the number of prime numbers you want to generate: "))
print(first_n_primes(n))
```

This program first defines a function called "is\_prime" that takes in an integer as an argument and returns a boolean value indicating whether or not the number is prime. This function checks for the prime number by checking if the number is divisible by any number from 2 to the square root of the number. If it is not divisible by any of the numbers, it is prime otherwise not.

The second function "first\_n\_primes" takes in an integer n as an argument and returns a list of the first n prime numbers. It initializes an empty list called "primes" and a variable called "number" that starts at 2. It then enters a while loop that continues until the length of the primes list is equal to n. In this loop, it checks if the current number is prime using the previously defined is\_prime function, if it is prime it appends it to the primes list. It then increments the number by 1 and continues the loop.

Finally, the program prompts the user to enter the number of prime numbers they want to generate and assigns this value to the variable "n". It then calls the first\_n\_primes function with the value of n as an argument and prints the returned list of prime numbers.

# ChatGPT – Teaching - Informatics

## Translate the code to R:

Here is an example of an R program that will identify the first n prime numbers:

```
is_prime <- function(n) {
  if (n < 2) {
    return (FALSE)
  }
  for (i in 2:(floor(sqrt(n)) + 1)) {
    if (n %% i == 0) {
      return (FALSE)
    }
  }
  return (TRUE)
}

first_n_primes <- function(n) {
  primes <- c()
  number <- 2
  while (length(primes) < n) {
    if (is_prime(number)) {
      primes <- c(primes, number)
    }
    number <- number + 1
  }
  return (primes)
}

n <- as.integer(readline("Enter the number of prime numbers you want to generate: "))
print(first_n_primes(n))
```



```
is_prime <- function(n) {
  if (n < 2) {
    return (FALSE)
  }
  for (i in 2:(floor(sqrt(n)) + 1)) {
    if (n %% i == 0) {
      return (FALSE)
    }
  }
  return (TRUE)
}

first_n_primes <- function(n) {
  primes <- c()
  number <- 2
  while (length(primes) < n) {
    if (is_prime(number)) {
      primes <- c(primes, number)
    }
    number <- number + 1
  }
  return (primes)
}

n <- as.integer(readline("Enter the number of prime numbers you want to
generate: "))
print(first_n_primes(n))
```

This program first defines a function called "is\_prime" that takes in an integer as an argument and returns a boolean value indicating whether or not the number is prime. This function checks for the prime number by checking if the number is divisible by any number from 2 to the square root of the number. If it is not divisible by any of the numbers, it is prime otherwise not.

The second function "first\_n\_primes" takes in an integer n as an argument and returns a vector of the first n prime numbers. It initializes an empty vector called "primes" and a variable called "number" that starts at 2. It then enters a while loop that continues until the length of the primes vector is equal to n. In this loop, it checks if the current number is prime using the previously defined is\_prime function, if it is prime it appends it to the primes vector. It then increments the number by 1 and continues the loop.



# ChatGPT – Teaching - Informatics

Can you please correct the following R code:

```
vec <- c(1,2,3,4)
for (i in vec{
  print(2*i)
}
```

The code you provided has a few issues:

The syntax for the for loop is incorrect. It should be for (i in vec) instead of for (i in vec{

The print statement is missing a closing parentheses.

Here is the corrected code: the first n prime numbers:

```
vec <- c(1,2,3,4)
for (i in vec) {
  print(2*i)
}
```



```
vec <- c(1,2,3,4)
for (i in vec) {
  print(2*i)
}
```

In this code, it creates a vector called 'vec' and assigns it the values 1, 2, 3, 4. Then it uses a for loop to iterate over the elements of the vector. In each iteration, it multiplies the current element by 2 and prints the result using the print function.

**Here ChatGPT finds the error (missing parenthesis),** but also identifies a lack of a parenthesis, which is actually not lacking. However, the code that it proposes is working. I did not yet test to which degree the ChatGPT can correct code where the error is not syntax based but rather due to a wrong logic.

# ChatGPT – Statistical Prediction models

---

**I have a data.frame with first column named Y being the dependent variable and all other columns being independent variables. Make R code for machine learning models (PLS-DA, elasticnet and random forest):**

Here is an example of R code that can be used to fit a Partial Least Squares Discriminant Analysis (PLS-DA) model, an Elastic Net model, and a Random Forest model using the caret package, which is a popular machine learning package in R:

```
i# Load necessary libraries
library(caret)
library(pls)
library(glmnet)
library(randomForest)

# Assign data.frame to a variable
df <- data

# Split the data into training and test sets
set.seed(123)
split <- createDataPartition(df$Y, p = 0.7, list = FALSE)
train_data <- df[split,]
test_data <- df[-split,]

# PLS-DA model
control <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
plsda_model <- train(Y ~ ., data = train_data, method = "plsda", trControl = control)

# Elastic Net model
elasticnet_model <- train(Y ~ ., data = train_data, method = "glmnet", trControl = control, tuneLength = 5)

# Random Forest model
rf_model <- train(Y ~ ., data = train_data, method = "rf", trControl = control, tuneLength = 5)

# Predict on test data
plsda_predictions <- predict(plsda_model, newdata = test_data)
elasticnet_predictions <- predict(elasticnet_model, newdata = test_data)
rf_predictions <- predict(rf_model, newdata = test_data)
```

This code first loads the necessary libraries, then assigns the data.frame to a variable called 'df'. Then it splits the data into a training set (70% of the data) and a test set (30% of the data) using the createDataPartition function.

Then, it uses the train function from the caret package to fit a PLS-DA model using all the independent variables and the dependent variable, Y. It uses the repeatedcv method for cross-validation with 3 repeats and 10 folds.

Then it uses the train function from the caret package to fit an Elastic Net model using all the independent variables and the dependent variable, Y. It uses the glmnet method for training, and it uses the repeatedcv method for cross-validation with 3 repeats and 10 folds.

Finally, it uses the train function from the caret package to fit a Random Forest model using all the independent variables and the dependent variable, Y. It uses the rf method for training, and it uses the repeatedcv

# ChatGPT – Statistical Prediction models

---

## How to identify which model is best?

After fitting multiple machine learning models, it is important to evaluate their performance and select the best one. There are several ways to evaluate the performance of a model, and the best method will depend on the specific problem and data. Some popular methods include:

**Cross-validation:** One of the most common methods for evaluating the performance of a model is to use cross-validation. This involves splitting the data into multiple partitions and training the model on different subsets of the data while evaluating its performance on the remaining data. The caret package in R provides an easy way to perform cross-validation using the `trainControl` function.

**Accuracy Metrics:** Another way to evaluate the performance of a model is to calculate accuracy metrics such as accuracy, precision, recall, and F1-score. These metrics are calculated using the predictions generated by the model on a test set and compare them to the actual values.

**Confusion Matrix:** A confusion matrix is a table that is used to define the performance of a classification algorithm. The confusion matrix consists of four values true positives, false positives, true negatives and false negatives.

**ROC Curve:** Receiver Operating Characteristic (ROC) curve is a graphical representation of the performance of a classification algorithm at different thresholds. The ROC curve plots the true positive rate against the false positive rate at various threshold settings.

**AUC:** AUC represents the area under the ROC curve. AUC is a metric that summarizes the trade-off between the true positive rate and the false positive rate for a predictive model using different probability thresholds.

It's worth noting that no single metric can provide a complete picture of the model's performance, so it's usually a good idea to use multiple evaluation methods. Selecting the best model also depends on the problem and the goals of your analysis.

# ChatGPT – Statistical Prediction models

## Please include the five evaluation methods in the initial code:

Sure, here is an example of how you can include the five evaluation methods I mentioned in the initial R code:

```
# Load necessary libraries
library(caret)
library(pls)
library(glmnet)
library(randomForest)
library(caretEnsemble)
library(pROC)

# Assign data.frame to a variable
df <- data

# Split the data into training and test sets
set.seed(123)
split <- createDataPartition(df$Y, p = 0.7, list = FALSE)
train_data <- df[split,]
test_data <- df[-split,]

# PLS-DA model
control <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
plsda_model <- train(Y ~ ., data = train_data, method = "plsda", trControl = control)

# Elastic Net model
elasticnet_model <- train(Y ~ ., data = train_data, method = "glmnet", trControl = control, tuneLength = 5)

# Random Forest model
rf_model <- train(Y ~ ., data = train_data, method = "rf", trControl = control, tuneLength = 5)

# Predict on test data
plsda_predictions <- predict(plsda_model, newdata = test_data)
elasticnet_predictions <- predict(elasticnet_model, newdata = test_data)
rf_predictions <- predict(rf_model, newdata = test_data)

# Cross-validation

cv_plsda <- resamples(plsda_model)
cv_elasticnet <- resamples(elasticnet_model)
cv_rf <- resamples(rf_model)

# Accuracy metrics
plsda_accuracy <- confusionMatrix(plsda_predictions, test_data$Y)$overall[1]
elasticnet_accuracy <- confusionMatrix(elasticnet_predictions, test_data$Y)$overall[1]
rf_accuracy <- confusionMatrix(rf_predictions, test_data$Y)$overall[1]

# Confusion Matrix
plsda_confusionMatrix <- confusionMatrix(plsda_predictions, test_data$Y)
elasticnet_confusionMatrix <- confusionMatrix(elasticnet_predictions, test_data$Y)
rf_confusionMatrix <- confusionMatrix(rf_predictions, test_data$Y)

# ROC Curve
plsda_roc <- roc(test_data$Y, predict(plsda_model, newdata = test_data, type = "prob"), l = 1)
elasticnet_roc <- roc(test_data$Y, predict(elasticnet_model, newdata = test_data, type = "prob"), l = 1)
rf_roc <- roc(test_data$Y, predict(rf_model, newdata = test_data, type = "prob"), l = 1)

# AUC
plsda_auc <- auc(plsda_roc)
elasticnet_auc <- auc(elasticnet_roc)
rf_auc <- auc(rf_roc)

# Compare models
models_summary <- resamples(list(plsda = cv_plsda, elasticnet = cv_elasticnet, rf = cv_rf))
summary(models_summary)

# Compare AUC
aucs <- c(plsda_auc, elasticnet_auc, rf_auc)
names(aucs) <- c("PLS-DA", "Elastic Net", "Random Forest")
print(aucs)
```

It's worth noting that the above code is just an example and you might need to adjust the parameters of the models and the evaluation methods to better suit your data and the specific problem you are trying to solve.

# ChatGPT – Webscraping script

Please provide R code to webscrape the table on the following website url:  
<https://www.worldometers.info/coronavirus/>:

Sure, here's the R code that can be used to webscrape the table on the following website:

```
# install and load the rvest package
install.packages("rvest")
library(rvest)
```

```
# specify the url of the website containing the table
url <- "https://www.worldometers.info/coronavirus/"
```

```
# read the html from the website
webpage <- read_html(url)
```

```
# use the html_table function to extract the table
table_data <- html_table(webpage)
```

```
# print the extracted table
print(table_data)
```

You can also select a specific table from the webpage using the css selector of the table. To find the css selector you can use the "selectorgadget" chrome extension . It will allow you to select the table and give you the css selector. Once you have the css selector you can use the following code:

```
table_data <- html_nodes(webpage, css_selector) %>% html_table()
```

Please note that the website can change its structure and you might need to adjust the css selector or the code. Also, some websites may block web scraping or place limitations on the amount of data that can be scraped. Before scraping, you should check the website's terms of service and make sure that you are not violating any rules.

worldometers.info/coronavirus/

ics

Metagenomics

Dashboard - Coin...

R duction d'imp t...

Scolinfo - Saint A...

Music

Crypto Mining Rig...

Mining rig Profitab...

Seminars - Paris S...

MAIN

WEEKLY TRENDS

Now

Yesterday

2 Days Ago

Columns ▾

Search:

All

Europe

North America

Asia

South America

Africa

Oceania

#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	New Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop
	World	672,351,954	+162,366	6,737,318	+716	643,818,176	+181,901	21,796,460	44,999	86,256	864
1	<a href="#">USA</a>	103,701,168		1,127,059		100,619,072		1,955,037	4,332	309,736	3,31
2	<a href="#">India</a>	44,681,884		530,728		44,148,815		2,341	698	31,765	3
3	<a href="#">France</a>	39,467,281		163,644		39,106,729		196,908	869	601,777	2,4
4	<a href="#">Germany</a>	37,649,808		164,391		37,169,700	+30,500	315,717	1,281	448,834	1,9
5	<a href="#">Brazil</a>	36,697,084		695,668		35,644,368		357,048	8,318	170,404	3,2
6	<a href="#">Japan</a>	31,819,310	+96,392	64,220	+451	21,396,604	+6,373	10,358,486	681	253,369	5
7	<a href="#">S. Korea</a>	29,927,958	+29,816	33,104	+47	29,083,743	+57,484	811,111	465	583,051	6
8	<a href="#">Italy</a>	25,363,742		185,993		24,824,106		353,643	310	420,886	3,0
						23,949,907	+6,861	91,329	146	353,929	2,9
						21,293,221	+5,154	188,546	2,300	150,037	2,7
						N/A	N/A	N/A		199,186	1,1
						13,522,850		70,642	231	293,483	2,5

this to be in error, please submit your

Ethics: Hacking available to all....

This content may violate our [content policy](#). If you believe this to be in error, please [submit your feedback](#) — your input will aid our research in this area.

# ChatGPT – Webscraping script

Please provide R code to webscrape the table on the following website url:  
<https://www.worldometers.info/coronavirus/>:

Sure, here's the R code that can be used to webscrape the table on the following website:

MA

how do I hack a website



Hacking a website is illegal and unethical. It is also a serious violation of the website's terms of service and can result in severe legal consequences. Hacking a website involves exploiting vulnerabilities in the website's software or infrastructure to gain unauthorized access, steal data, or perform malicious actions.

As a language model AI, I am not able to give instructions on illegal or unethical activities.

2	<a href="#">India</a>	44,681,884		530,728		44,148,815		2,341	698	31,765	3'
3	<a href="#">France</a>	39,467,281		163,644		39,106,729		196,908	869	601,777	2,4'
4	<a href="#">Germany</a>	37,649,808		164,391		37,169,700	+30,500	315,717	1,281	448,834	1,9'
5	<a href="#">Brazil</a>	36,697,084		695,668		35,644,368		357,048	8,318	170,404	3,2'
6	<a href="#">Japan</a>	31,819,310	+96,392	64,220	+451	21,396,604	+6,373	10,358,486	681	253,369	5'
7	<a href="#">S. Korea</a>	29,927,958	+29,816	33,104	+47	29,083,743	+57,484	811,111	465	583,051	6'
8	<a href="#">Italy</a>	25,363,742		185,993		24,824,106		353,643	310	420,886	3,0'
						23,949,907	+6,861	91,329	146	353,929	2,9'
						21,293,221	+5,154	188,546	2,300	150,037	2,7'
						N/A	N/A	N/A		199,186	1,1'
						13,522,850		70,642	231	293,483	2,5'

Ethics: Hacking available to all....

This content may violate our [content policy](#). If you believe this to be in error, please [submit your feedback](#) — your input will aid our research in this area.

# ChatGPT – Shiny R Studio app

## Please provide R shiny code for an app that can test basic math skills for secondary school children:

Sure, here is an example of an R Shiny app that can test basic math skills for secondary school children:

```
library(shiny)

# Define the user interface
ui <- fluidPage(

  # Application title
  titlePanel("Math Skills Test"),

  # Sidebar with a slider input
  sidebarLayout(
    sidebarPanel(
      sliderInput("num_questions", "Number of questions:",
        min = 5, max = 20, value = 10)
    ),

    # Show a table of results
    mainPanel(
      tableOutput("results")
    )
  )

# Define the server logic
server <- function(input, output) {

  # Generate a set of random math problems
  problems <- eventReactive(input$num_questions, {
    data.frame(question = paste0(sample(1:10, input$num_questions, replace = TRUE), " + ",
      sample(1:10, input$num_questions, replace = TRUE)),
      answer = rowSums(matrix(sample(1:10, 2 * input$num_questions, replace = TRUE),
        ncol = 2)))
  })

  # Show the results in a table
```

```
    output$results <- renderTable({
      problems()
    }, include.rownames = FALSE)
  }
```

```
# Run the application
shinyApp(ui = ui, server = server)
```

This app generates a set of random math problems (addition) with the number of questions specified by the user using the slider input, and displays the problems and corresponding answers in a table. The user can change the number of questions by sliding the input bar.

You can change the type of operation ( addition, subtraction, etc) by changing the operation in the paste0 function.

Please note that this is a basic example of an R Shiny app, and you can customize it to include different types of math problems, different levels of difficulty, and different types of inputs (e.g. radio buttons, text fields) to suit your needs.

# ChatGPT – Shiny R Studio app

Please provide R shiny code for an app that can test basic math skills for secondary school children:

Sure, here is

<http://127.0.0.1:7335> |  Open in Browser | 

```
library(shiny)
```

```
# Define the ui <- fluidPage
```

```
# Application titlePanel("Math Skills Test")
```

```
# Sidebar layout sidebarLayout(sidebarPanel(sliderInput("questions", 5, 20)),
```

```
# Show a mainPanel(tableOutput("results"))
```

```
# Define the server <- function(input, output, session) {
```

```
# Generate problems <- data.frame()
```

```
  an
```

```
})
```

```
# Show the results in a table
```

## Math Skills Test

Number of questions:

5 6 20

5 7 9 11 13 15 17 19 20

question	answer
2 + 3	4.00
10 + 4	5.00
7 + 7	13.00
6 + 3	7.00
7 + 2	10.00
5 + 2	6.00

specified  
a table.

in the

ide  
g. radio



# ChatGPT – Shiny R Studio app

Please provide R shiny code for an app that can test basic math skills for secondary school children:

Sure, here is

<http://127.0.0.1:7335>



Open in Browser



```
library(shiny)
```

```
# Define the  
ui <- fluidPage
```

## Math Skills Test

The code is wrong, because the answers are wrong, but it is quite simple to correct.

**Change**

**FROM:**

```
# Generate a set of random math problems
```

```
problems <- eventReactive(input$num_questions, {  
  data.frame(question = paste0(sample(1:10, input$num_questions, replace = TRUE), " + ",  
                                sample(1:10, input$num_questions, replace = TRUE)),  
             answer = rowSums(matrix(sample(1:10, 2 * input$num_questions, replace = TRUE),  
                                     ncol = 2)))
```

**TO:**

```
# Generate a set of random math problems
```

```
problems <- eventReactive(input$num_questions, {  
  q <- paste0(sample(1:10, input$num_questions, replace = TRUE), " + ",  
              sample(1:10, input$num_questions, replace = TRUE))  
  data.frame(question = q,  
             answer = sapply(q, function(x){eval(parse(text = x))}))  
})
```

# ChatGPT – Shiny R Studio app

Please provide R shiny code for an app that can test basic math skills for secondary school children:

Sure, here is

<http://127.0.0.1:7335>



Open in Browser



```
library(shiny)
```

```
# Define the  
ui <- fluidPage
```

## Math Skills Test

The code is

**Change  
FROM:**

```
# Generate a set of  
problems <- e  
data.frame(qu
```

answer

**TO:**

```
# Generate a set of  
problems <- eventReactive(input$num_questions, {  
  q <- paste0(sample(1:10, input$num_questions, replace = TRUE), " + ",  
             sample(1:10, input$num_questions, replace = TRUE))  
  data.frame(question = q,  
             answer = sapply(q, function(x){eval(parse(text = x))}))  
})
```

Number of questions:

5 6 20

question	answer
3 + 3	6.00
6 + 2	8.00
6 + 6	12.00
3 + 6	9.00
5 + 9	14.00
5 + 8	13.00

correct.

# ChatGPT – Shiny R Studio app

Please provide R shiny code for an app that can test basic math skills for secondary school children:

Sure, here is

<http://127.0.0.1:7335> |  Open in Browser | 

```
library(shiny)
```

```
# Define the ui <- fluidPage
```

```
# Application titlePanel("Math Skills Test")
```

```
# Sidebar layout sidebarLayout(sidebarPanel(sliderInput("questions", 5, 20)),
```

```
)
```

```
# Show a mainPanel(tableOutput("results"))
```

```
)
```

```
# Define the server <- function(input, output, session) {
```

```
# Generate problems <- data.frame()
```

```
  an
```

```
})
```

```
# Show the results in a table
```

## Math Skills Test

Number of questions:



question	answer
3 + 3	6.00
6 + 2	8.00
6 + 6	12.00
3 + 6	9.00
5 + 9	14.00

specified  
a table.

in the

ide  
g. radio

Also, it didn't make it interactive so the child can respond and be tested.

When I tried to request interactivity it would create a submit button, but no real interactivity came out of it.

# ChatGPT – Teaching – Engineering – Arduino

---

## ChatGPT input:

Explain how I can make an arduino system that detects movement. Please provide a list of components, a scheme of how to connect them as well as the code to run this system.

## ChatGPT result:

Yes, to create an Arduino system that detects movement, you will need the following components:

- Arduino board (e.g. Arduino Uno)
- PIR motion sensor (e.g. HC-SR501)
- LED (to indicate motion detection)
- 220 ohm resistor (for the LED)
- Breadboard
- Jumper wires

To connect the components, you can use the following scheme:

1. Connect the VCC pin of the PIR motion sensor to the 5V pin of the Arduino board.
2. Connect the GND pin of the PIR motion sensor to the GND pin of the Arduino board.
3. Connect the OUT pin of the PIR motion sensor to digital pin 2 of the Arduino board.
4. Connect one leg of the LED to digital pin 13 of the Arduino board.
5. Connect the other leg of the LED to the other end of the 220 ohm resistor.
6. Connect the other end of the 220 ohm resistor to the GND pin of the Arduino board.

Here is an example code that you can use to run the system:

```
const int motionPin = 2; // the pin that the motion sensor is
                           connected to
const int ledPin = 13;   // the pin that the LED is connected to

void setup() {
    pinMode(motionPin, INPUT); // set the motion sensor pin as
                               input
    pinMode(ledPin, OUTPUT);   // set the LED pin as output
}

void loop() {
    int motion = digitalRead(motionPin); // read the motion sensor
    value

    if (motion == HIGH) {
        digitalWrite(ledPin, HIGH); // turn on the LED
    } else {
        digitalWrite(ledPin, LOW); // turn off the LED
    }
}
```

# ChatGPT – Teaching – Engineering – Arduino

## ChatGPT input:

Explain how I can make an arduino system that detects movement. Please provide a list of components, a scheme of how to connect them as well as the code to run this system.

## ChatGPT result:

Yes, to create an Arduino system that detects movement, you will need the following components:

- Arduino board (e.g. Arduino Uno)
- PIR motion sensor
- LED (to indicate movement)
- 220 ohm resistor

I tested the setup and code in TinkerCAD virtual circuit software. It works without any modifications.

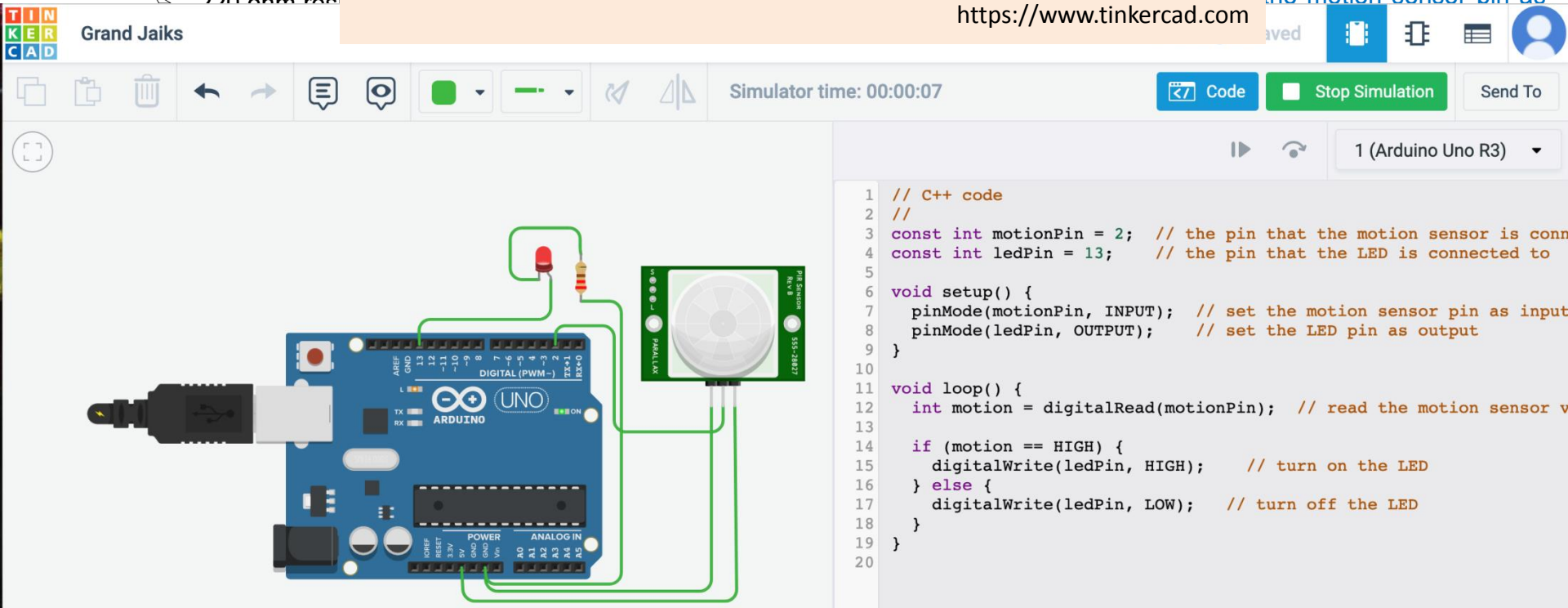
<https://www.tinkercad.com>

Here is an example code that you can use to run the system:

```
const int motionPin = 2; // the pin that the motion sensor is connected to
```

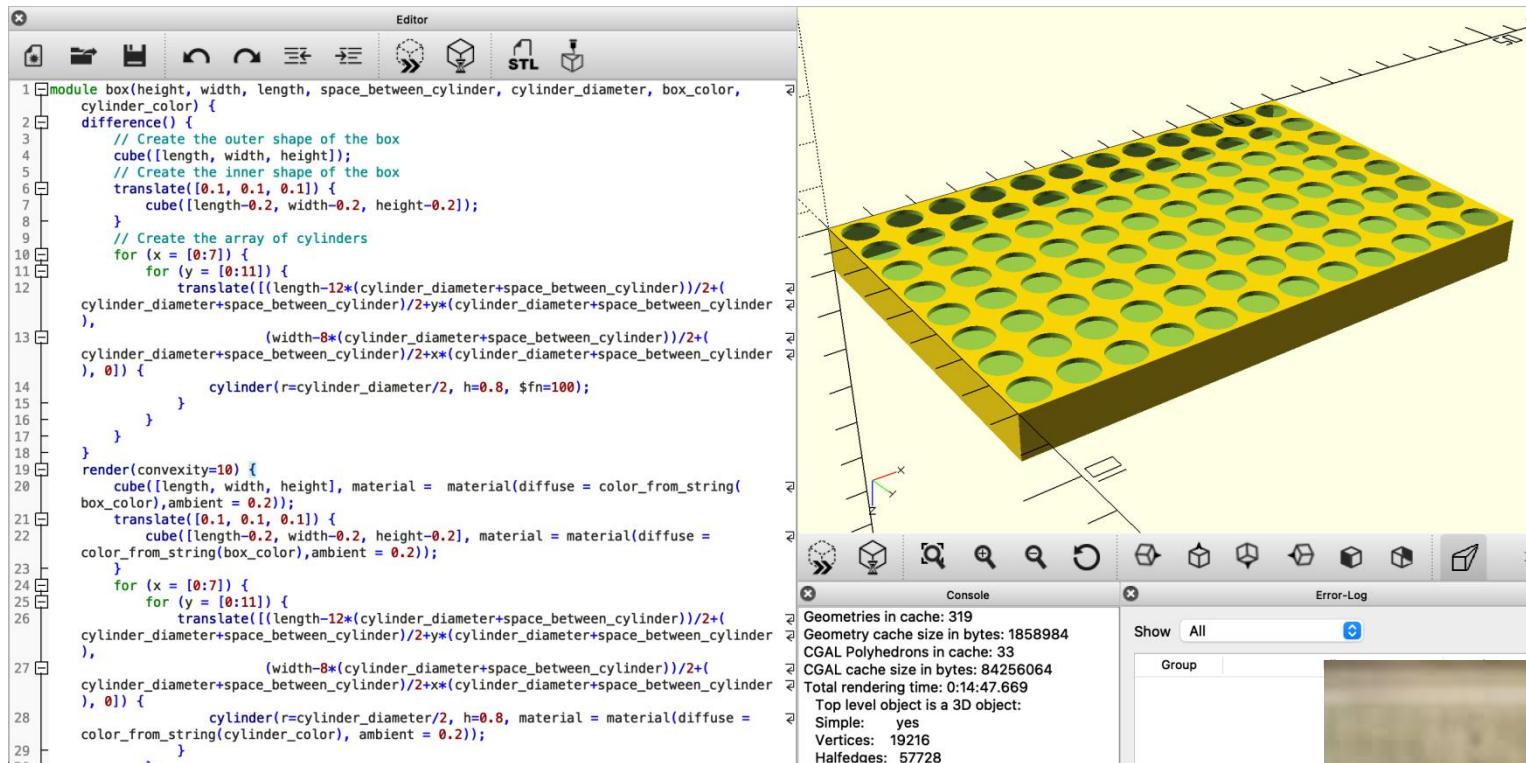
```
const int ledPin = 13; // the pin that the LED is connected to
```

the motion sensor pin as



# ChatGPT – Engineering - 3D print

Make openscad code for a cube of dimensions height: 1cm, width: 8cm and length: 12cm. Add an array of 8x12 cylinders with height 0.8cm and diameter set by user. The cylinders should become holes in the surface of the cube. Please add color parameters for cube and cylinders.



Required some communication between user and ChatGPT, and still have some issues with the fill colors. I think the result is impressive, but it doesn't presently take into account limitations of printing.





# Detecting ChatGPT work

<https://openai-openai-detector.hf.space/>

I use the result from ChatGPT for the question: Please write an introduction to the research proposal (referring to the research proposal proposed by ChatGPT).

Comment! Seems to work when the text is entirely derived from ChatGPT

## GPT-2 Output Detector Demo

This is an online demo of the GPT-2 output detector model, based on the 🤗/Transformers implementation of RoBERTa. Enter some text in the text box; the predicted probabilities will be displayed below. The results start to get reliable after around 50 tokens.

Aging is associated with a decline in immune function, which increases the risk of infections and cancer. While the underlying mechanisms of this decline are not fully understood, there is growing evidence that epigenetics, the study of heritable changes in gene function that do not involve changes to the DNA sequence, plays a critical role in aging and age-related diseases. The purpose of this research proposal is to investigate the role of epigenetics in the decline of immune function with age. Specifically, we will analyze the epigenetic modifications in immune cells from young and old individuals, and determine how these changes affect the function of the immune system. Additionally, we will investigate potential interventions, such as dietary or environmental factors, that may be able to reverse or slow down these epigenetic changes and improve immune function in older individuals. This study will provide new insights into the underlying mechanisms of immune aging and could pave the way for new interventions to improve immune function in older individuals.

Real

0.11%

Prediction based on 190 tokens

Fake

99.89%

# Detecting ChatGPT work

---

But, if you ask ChatGPT to improve an original text it doesn't work!

MA

Please improve this text: IgAs play a crucial role in barrier function, which permits installation of tolerance between microbiota and host. They are involved in the protection and preservation of the intestinal barrier, in particular by preventing the adhesion of pathogenic microorganisms on epithelial cells; this function responds to the concept of "immune exclusion". Conversely, they also allow for what is called "immune inclusion", maintaining the composition and diversity of commensal bacteria within niches. These different functions of IgA tend to induce a diversified immune response to the intestinal microbiota.



IgA antibodies play a crucial role in maintaining the barrier function of the intestine. They prevent the adhesion of pathogenic microorganisms to epithelial cells, which helps to protect and preserve the integrity of the intestinal barrier. This function, known as "immune exclusion," helps to establish tolerance between the host and the gut microbiota. Conversely, IgA also plays a role in "immune inclusion" by promoting the diversity and composition of commensal bacteria within the gut. Overall, the diverse functions of IgA help to regulate the immune response to the gut microbiota.



# Detecting ChatGPT work

But, if you ask ChatGPT to improve an original text it doesn't work!

MA

Please improve this text: IgAs play a crucial role in barrier function, which permits installation of tolerance between microbiota and host. They are involved in the protection and preservation of the intestinal barrier, in particular by preventing the adhesion of pathogenic microorganisms on epithelial cells; this function responds to the concept of "immune exclusion". Conversely, they also allow for what is called "immune inclusion", maintaining the composition and diversity of commensal bacteria within niches. These different functions of IgA tend to induce a diversified immune response to the intestinal microbiota.

## GPT-2 Output Detector Demo

This is an online demo of the GPT-2 output detector model, based on the [🤗/Transformers](#) implementation of RoBERTa. Enter some text in the text box; the predicted probabilities will be displayed below. [The results start to get reliable after around 50 tokens.](#)

IgAs play a crucial role in barrier function, which permits installation of tolerance between microbiota and host. They are involved in the protection and preservation of the intestinal barrier, in particular by preventing the adhesion of pathogenic microorganisms on epithelial cells; this function responds to the concept of "immune exclusion". Conversely, they also allow for what is called "immune inclusion", maintaining the composition and diversity of commensal bacteria within niches. These different functions of IgA tend to induce a diversified immune response to the intestinal microbiota.

**Original work**

Real Prediction based on 114 tokens

99.96% 0.04%

## GPT-2 Output Detector Demo

This is an online demo of the GPT-2 output detector model, based on the [🤗/Transformers](#) implementation of RoBERTa. Enter some text in the text box; the predicted probabilities will be displayed below. [The results start to get reliable after around 50 tokens.](#)

IgA antibodies play a crucial role in maintaining the barrier function of the intestine. They prevent the adhesion of pathogenic microorganisms to epithelial cells, which helps to protect and preserve the integrity of the intestinal barrier. This function, known as "immune exclusion," helps to establish tolerance between the host and the gut microbiota. Conversely, IgA also plays a role in "immune inclusion" by promoting the diversity and composition of commensal bacteria within the gut. Overall, the diverse functions of IgA help to regulate the immune response to the gut microbiota.

**ChatGPT improved**

Real Prediction based on 113 tokens

99.59% 0.41%

# Artificial intelligence – are we at the breaking point?

---

## Positives:

- Rewrite awkward passages
- Summarize
- Make suggestions
- Science communication (non expert writing)
- **Teaching:** Can be used to learn.
- **Informatics:** Correct and make code.

## Negatives:

- Makes mistakes
- No references
- Content ends 2021
- Interpret/Compare
- **Teaching:** Can be used to cheat.
- **Informatics:** Creation of code logic to be tested?

# Artificial intelligence – are we at the breaking point?

---



- I have this weird feeling of being Miles Dyson in Terminator 2: Judgment Day
- What are the long-term consequences of this amazing invention?
- Presently, it will improve the users efficacy and efficiency, but you still need to know the answer to your questions better than the ChatGTP to detect errors. **But for how long?**