

Using Generative AI as a tool

What are generative AI models?

- Algorithms that have been trained by looking at millions of documents to recognize patterns
- These can then use this pattern recognition to answer a range of questions
- **Is this an opportunity or threat?**

What are the major large language models

- ChatGPT
- **Claude**
- PaLM
- BERT
- Falcon
- And many others....

What are the potential uses for writing?

- **Generate a basic summary of a topic**
 - Developing a summary to help generate ideas
- **Understand the important Sources surrounding an important topic**
 - Gain an understanding of where key sources exist
 - Obtain Summaries of these sources
- **Outline and Structure an Argument**
 - present information helpful to formulating an argument or an explanation for a stated position.
 - Draft initial text on a topic
- **Check Grammar and Structure of written work**
 - Iterate drafts to improve readability and grammar

Generating a summary

- Step 1: Find Your Article
- Step 2: Prepare Your Article
- Step 3: Ask for a Summary
- Command

please summarize the article at

<https://www.belfercenter.org/publication/seeing-data-structure>

Get a list of important sources

Prompt: Give me a list of ten primary sources on "The Maize nested association mapping population"?

Here are ten primary sources on the Maize Nested Association Mapping (NAM) population:

1. McMullen, M. D., et al. (2009). "Genetic Properties of the Maize Nested Association Mapping Population." *Science*. This foundational paper discusses the creation and genetic structure of the NAM population, highlighting its utility for high-resolution quantitative trait locus (QTL) mapping.
2. Gage, J. L., et al. (2020). "Ten Years of the Maize Nested Association Mapping Population: Impact, Limitations, and Future Directions." *Genetics*. This review covers the impact and contributions of the NAM population to maize genetics and breeding over a decade, identifying both strengths and areas for improvement.
3. Zhang, N., et al. (2015). "Genome-Wide Association of Carbon and Nitrogen Metabolism in the Maize Nested Association Mapping Population." *Plant Physiology*. This study uses the NAM population to explore the genetic basis of carbon and nitrogen metabolism in maize, linking these traits to crop yield and adaptation.
4. Cook, J. P., et al. (2012). "Genetic Architecture of Maize Kernel Composition in the Nested Association Mapping and Inbred Association Panels." *Plant Physiology*. This research investigates the genetic determinants of maize kernel composition using the NAM population, providing insights into nutritional quality and breeding.

Generating an outline

Prompt: Generate an outline for a 5 page paper on "How well does the Maize nested association mapping population represent the diversity of global maize"?

```
##### I. Introduction
- **Background on Maize Diversity**
  - Importance of maize as a global staple crop
  - Genetic diversity and its role in crop improvement
- **Introduction to the Maize Nested Association Mapping (NAM) Population**
  - Overview of NAM population development
  - Purpose and significance in genetic studies
- **Thesis Statement**
  - This paper evaluates how well the NAM population represents the genetic diversity of global maize varieties.

##### II. The Genetic Structure of the NAM Population
- **Development and Design**
  - Description of the creation of the NAM population
  - Parental lines used and rationale for their selection
- **Genetic Properties and Composition**
  - Analysis of the genetic diversity within the NAM population
  - Comparison to natural maize populations

##### III. Representativeness of Global Maize Diversity
- **Geographic and Genetic Representation**
  - Assessment of geographic origins of parental lines
  - Extent to which global maize genetic diversity is covered
- **Phenotypic Diversity**
  - Examination of the phenotypic traits represented in the NAM population
  - Comparison with traits found in diverse global maize varieties
- **Case Studies and Examples**
  - Specific examples of traits or genetic loci studied using the NAM population
  - Relevance to global maize diversity

##### IV. Applications and Limitations
- **Applications in Breeding and Research**
  - Success stories and contributions to maize genetics and breeding
  - Use of NAM population in identifying QTLs and understanding complex traits
- **Limitations and Gaps**
  - Critique of the limitations in representing certain regions or traits
  - Potential biases introduced by the choice of parental lines
- **Comparisons with Other Mapping Populations**
  - How the NAM population compares with other mapping populations in representing global diversity

##### V. Enhancing the Representativeness
- **Strategies for Improvement**
  - Suggestions for expanding the genetic base of the NAM population
  - Incorporation of additional diverse lines from underrepresented regions
- **Future Research Directions**
  - Areas for further study to enhance the utility of the NAM population
  - Emerging technologies and methodologies in genetic mapping and diversity studies

##### VI. Conclusion
- **Summary of Key Points**
  - Recap of the evaluation of the NAM population's representativeness
  - Implications for Maize Research and Breeding
  - The importance of genetic diversity in crop improvement
  - Final Thoughts
  - The NAM population as a valuable but evolving tool in global maize genetics

### References
```

Check Grammar

Prompt: Please correct the spelling and grammar of the following text. Show the corrections in bold so I can see what has been corrected.

Or

Prompt: Act as a proofreader and review the following text. Feel free to rephrase sentences or make changes to enhance clarity but maintain the overall tone and style of the original. Show all changes in bold so I can see what has been updated.

Use in Coding

- Basic Prompt
 - How to join two datasets in R?
 - How do I make a wide dataset long in R?
- More Complex Prompt
 - How can a create a ridgeline graph in ggplot using a colorblind friendly color scheme with added text for specific points with a star symbol

```
# Install and load necessary packages
install.packages("ggthemes")
install.packages("ggplot2")
install.packages("ggrepel")
install.packages("ggdist")

library(ggthemes)
library(ggplot2)
library(ggdist)
library(ggrepel)

# Prepare your data
set.seed(123)
data <- data.frame(
  x = rep(1:10, each = 100),
  y = rep(1:10, each = 100),
  group = rep(c("A", "B", "C"), each = 100)
)

# Create the ridgeline plot
p <- ggplot(data, aes(x = x, y = factor(y), fill = group)) +
  geom_density_ridges(scale = 0.9) +
  scale_fill_viridis(discrete = TRUE, option = "D") + # Colorblind-friendly palette
  theme_ridges() +
  theme(legend.position = "none")

# Annotation data
annotate_data <- data.frame(
  x = c(-2, 0, 2),
  y = c(3, 5, 7),
  label = c("Point A", "Point B", "Point C")
)

# Add annotations
p +
  geom_text_repel(data = annotate_data, aes(x = x, y = y, label = paste0("★ ", label)),
    color = "black", size = 5, nudge_y = 0.2)
```

Checking and debugging code

- Example Prompts from the chatGPT cheat sheet
 - Can you optimize the following R code?
 - Can you simplify the following R code?
 - Can you document the following code?
 - Can you check for readability of the following code?

Things to think about:

- Large language models do not understand the meaning of words
- Large language models prioritize standard American english
- Large language models do not protect user privacy

Other limitations

- Not Connected to the Internet
 - Models do not actually know things they are limited by what they have access to
- Hallucinating Facts and Sources
 - semantically or syntactically credible, but is actually false or nonsensical.
- Difficulty Long-form Content
 - May create content that is repetitive but inaccurate
- Biased Responses
 - The responses depend on the sources

Thoughts?