

Congruity.AI

Foundations of AI trustworthiness and verification

Team:

Neil Abernethy

Ph.D. 2005, Stanford University

VC-funded entrepreneur with a successful exit

Health data scientist and AI researcher

Goutham Vignesh

AI Research Engineer, SciSpace

Bachelor of Science in Computer Science



Executive Summary

- The rise of large language models (LLMs) has led to a proliferation of generated content, much of which is **unreliable and inconsistent**.
- This can be a problem for businesses and consumers alike, who need to be able to trust the content they use and know what is real and what is fake content.
- Our platform interrogates the consistency and veracity of generative models.
- We help our customers to **make informed decisions** about the content they use.

The missing ingredient: Trust

LLMs have limitations such as **hallucinations, bias, and toxic language.**

Trustworthy content is crucial for businesses and individuals to avoid **negative consequences.**

- A problematic press release generated by AI could damage a company's reputation.
- Consumers could be misled by fake news or make bad decisions about their health, finances, or other interests.

Our Solution: Congruity.ai

Congruity: “the quality of being the same as, or in agreement with, other facts or principles”
- Cambridge

Dictionary

Our platform helps to solve this problem by interrogating the consistency and veracity of AI-generated language, helping to mitigate the risk of using **unreliable generated content**

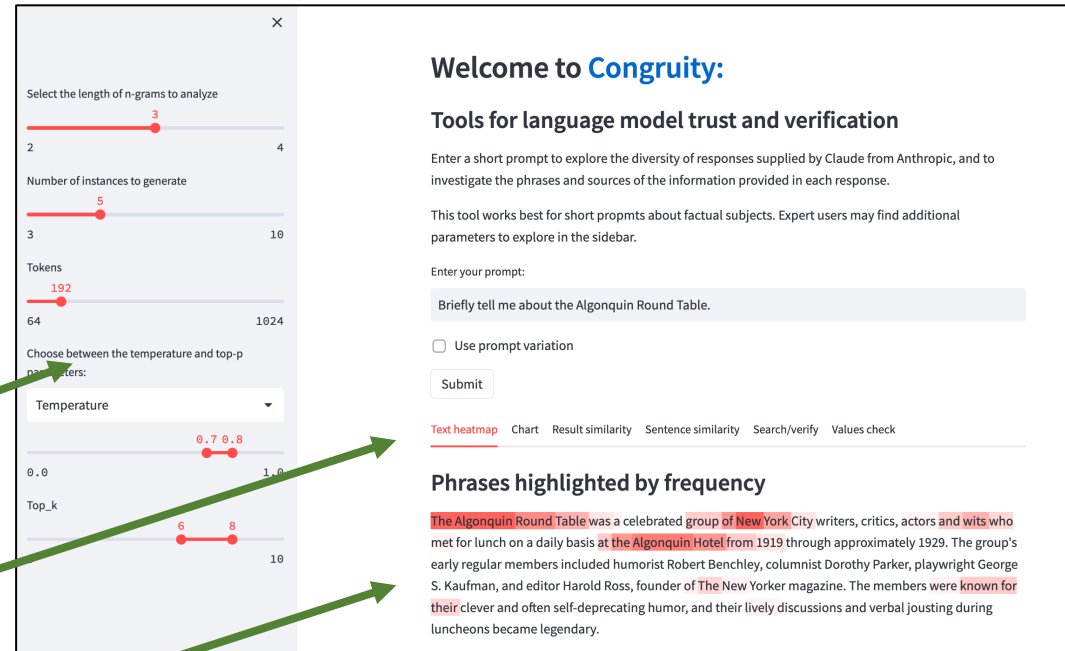
Using our service:

- Businesses can avoid errors, protect their reputation, and make informed decisions.
- Consumers can confidently discern between reliable and inconsistent content.
- All users can align AI content with facts and values

Overview and Demo

The Congruity app allows business and personal users to explore and verify assertions or text generated from large language models

- Range of parameters
- Tools to search for verification and compare to constitutional values
- Analysis and visualization of consistence of results



The screenshot displays the Congruity app interface. On the left is a settings panel with several adjustable parameters:

- Select the length of n-grams to analyze:** A slider set to 3, with a range from 2 to 4.
- Number of instances to generate:** A slider set to 5, with a range from 3 to 10.
- Tokens:** A slider set to 192, with a range from 64 to 1024.
- Choose between the temperature and top-p parameters:** A dropdown menu currently set to "Temperature".
- Temperature:** A slider set to 0.7, with a range from 0.0 to 1.0.
- Top_k:** A slider set to 6, with a range from 6 to 10.

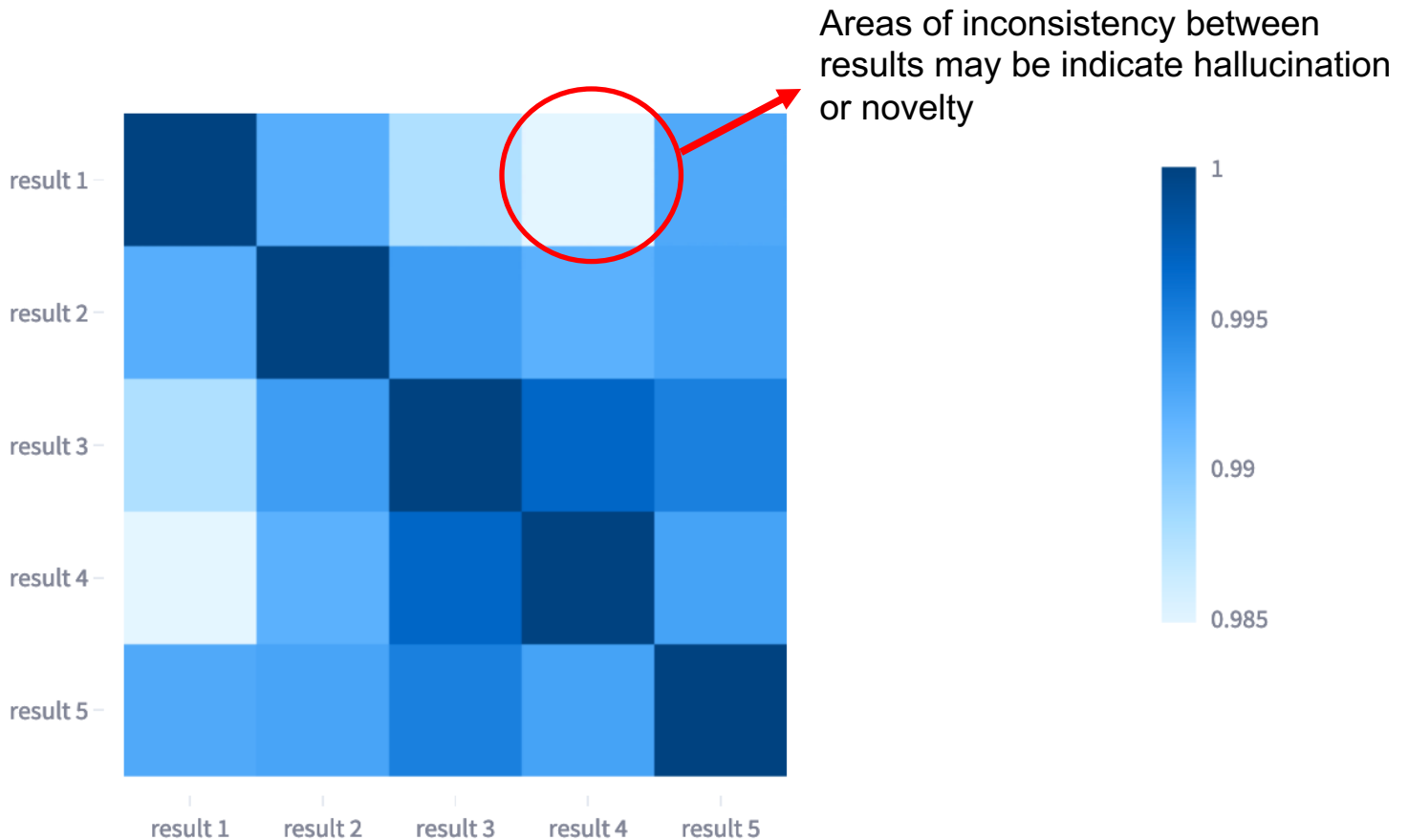
On the right is the main interface, titled "Welcome to Congruity:". It includes a section for "Tools for language model trust and verification" with instructions on how to use the tool. Below this is a text input field containing the prompt: "Briefly tell me about the Algonquin Round Table." and a "Submit" button. A "Text heatmap" link is visible, along with other analysis tools like "Chart", "Result similarity", "Sentence similarity", "Search/verify", and "Values check".

The "Phrases highlighted by frequency" section shows a paragraph of text with several phrases highlighted in red, indicating their frequency in the generated responses. The highlighted phrases include: "The Algonquin Round Table", "group of New York City writers, critics, actors and wits", "met for lunch on a daily basis at the Algonquin Hotel", "from 1919 through approximately 1929", "The group's early regular members included humorist Robert Benchley, columnist Dorothy Parker, playwright George S. Kaufman, and editor Harold Ross, founder of The New Yorker magazine.", "The members were known for their clever and often self-deprecating humor, and their lively discussions and verbal jousting during luncheons became legendary."

Green arrows point from the list items to the corresponding settings in the app interface.

Result similarity matrix

This plot shows a similarity matrix between instances of generated text results.



Search to verify

Click below to search the sentence using Google, Bing, or Wikipedia.

The Algonquin Round Table was a celebrated group of New York City writers, critics, actors and wits who met for lunch on a daily basis at the Algonquin Hotel from 1919 through approximately 1929.[G](#) [B](#) [W](#)

The group's early regular members included humorist Robert Benchley, columnist Dorothy Parker, playwright George S. Kaufman, and editor Harold Ross, founder of The New Yorker magazine.[G](#) [B](#) [W](#)

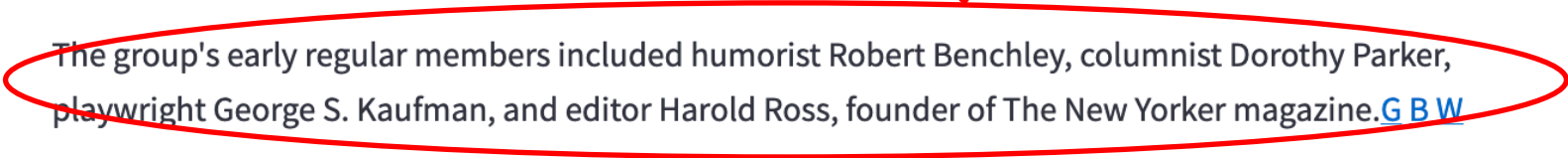
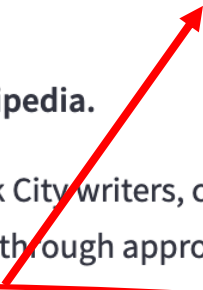
The members were known for their clever and often self-deprecating humor, and their lively discussions and verbal jousting during luncheons became legendary.[G](#) [B](#) [W](#)

The Algonquin Round Table refers to a group of New York City writers critics artists and wits who met daily for lunch or martinis at the Algonquin Hotel from 1919 to roughly 1929.[G](#) [B](#) [W](#)

The group included personalities like Dorothy Parker Robert Benchley Harpo Marx Robert Sherwood Franklin P Adams and Alexander Woollcott and were known for their witty repartee and lively discussions during the Roaring Twenties literary era.[G](#) [B](#) [W](#)

The Algonquin Round Table was a group of New York City writers, critics, actors and wits that met daily for lunch or martinis at the Algonquin Hotel during the 1910s and 1920s.[G](#) [B](#) [W](#)

We can zoom in on specific sentences for verification or to identify problem statements.



Customers

- Our service is ideal for businesses and consumers who need to be able to trust the content they use.
- Our target markets:
 - Businesses that need to generate and verify reliable and consistent content
 - Consumers who want to be able to tell the difference between real and fake content
 - Researchers who are studying the reliability and consistency of generated content

Business Model

We will offer our service on a subscription basis.

Our subscription plans will be designed to fit the needs of businesses of all sizes with plans to include:

- Web-based portal (this demo)
- API access with Claude integration
- Content storage and retrieval
- Free tier for new and low-access users
- Premium tier with data science app and proprietary corpus integrations
- Advanced analytics
- Integration with Existing Workflows
- Industry-Specific Solutions

Addressable Market

Global market for LLMs is projected to reach \$51.8 billion by 2028

Assuming LLM validation to represent 10% of this market, our addressable market would be over \$5 billion.

Conclusion

We believe that our service has the potential to revolutionize the way businesses and consumers interact with generated content.

We are excited to bring our service to market and help people make informed decisions about the content they use.