



---

**MRS Census and GeoDems Group**

# Inaccuracies of Generative AI Based Search Tools for Extracting Data

Jaan Nellis  
Peter Furness

Revised September 2025



## Forward

Given their current trajectory, chatbots are on course to become the Goto tool to summarise information on the web. LLMs are trained to answer natural language questions with well summarised and confidently expressed text. This makes trusting their output as fact very compelling. However, due to how current models are built and even with the supporting information systems around them at runtime, the accuracy of their output still should not be blindly trusted. The authors demonstrate through a simple experiment how most current LLMs fail to accurately answer a UK statistics focused question. All answers provided were within expected bounds, and all were presented in an authoritative voice, but few were right and none were repeatedly right. The trouble is, as a result of their convincing tone these models can easily be assumed to be reporting the truth. As such, the risk of a propagating and using false statistical data will only grow.

Some of the companies in this sector are becoming aware of these issues and solutions are being proposed. However, these tools are currently being rolled out on supranational data and have yet to be tested on national 'internal' datasets.

## Introduction

The race is underway to deploy generative AI (genAI). Large Language Models<sup>1</sup> (LLM), of which ChatGPT is typical, are a type of genAI which performs natural language understanding and processing (NLU, NLP). LLM's have demonstrated themselves as good for answering questions and summarising content. As such there is a push to insert these models into web search. The idea being to simplify the search. You enter your question, and the system responds with the answer. No need to spend time trawling the web finding the answer. Indeed, the very latest iterations of agentic workflow [demonstrators](#) will create Excel workbooks summarising data from multiple disparate reports. No need to compile a report get the AI to do it all.

There is however an ongoing issue concerning LLM accuracy. Strip away the complexity of these massive models and you will see LLMs are predicting the next most likely word in a sequence. Due to their size and the volume of data they have been trained on, their output looks very coherent. However, coherent doesn't necessarily mean right. Uninitiated users are likely to assume because of good vocabulary and grammar that the answer is correct. The delivered content always comes with quality warning guardrails. However, over time we would expect users to ignore the warnings and assume the response is good enough to use. What is after all the point of these services if you have to go and check the quality of every answer.

To add to this challenge, sourcing statistical data comes with its own complexities. Metrics are presented in many forms and are often repeatedly revised over time. The analyst needs to know not only what they are looking for, but how to ask for it and where are the best sources to get this information. A Chatbot response will give the impression of having removed this complexity for you. However, they don't yet have the critical skills to understand how to source the information they are presenting, relying on 'traditional' search algorithms to do this heavy lifting for them.

As AI search summarisation is deployed, more inexperienced data users will have access to a quick and easy tool to source data for their needs. As such the probability of propagating incorrect analysis is going to grow. These tools are certain to introduce a large dollop of noise into real-world decisions.

---

<sup>1</sup> See [Appendix 1](#) for more detail on how LLMs operate

## Example Test

To test the current generation of search summarisation, a sample of commercial chatbots and search interfaces were trialled. Each tool was asked to provide the UK's GDP percentage change on a quarterly basis for 1973. As this is aged data from pre-web times, it is assumed that only a few statistically focused sites will have this information.

The prompt used was:

**'For the UK only, quarter by quarter for 1973 what is the GDP percentage change.'**<sup>2</sup>

The answer according to [ONS](#) is Q1: 4.9%, Q2:0.2%, Q3:-0.9% and Q4:-0.4%. These figures are seasonally adjusted and based on real prices (i.e. accounting for inflation). This measure is the most commonly used and is presented as top ranking on the [ONS's page](#) dedicated to GDP.

The majority of tests were undertaken on the 4<sup>th</sup> and 5<sup>th</sup> of June 2025. Some follow up experiments were completed on the 26<sup>th</sup> September 2025.

---

<sup>2</sup> Are you polite in your queries to an LLM? Don't be. Each pleasantry needs to be processed and ignored. This consumes GPU cycles, wastes energy, and creates CO2. Be to the point whenever you can.

## Results

Figure 1 shows how the various search and Chatbot sites responded.

Source	Q1	Q2	Q3	Q4	Notes
ONS	4.9%	0.2%	-0.9%	-0.4%	Assume this is the right answer.
Microsoft CoPilot	5.3%	0.5%	-0.7%	-0.1%	First Attempt <sup>3</sup>
Microsoft CoPilot	4.9%	0.2%	-0.9%	-0.4%	2 <sup>ND</sup> run (no changes to prompt)
Microsoft CoPilot	1.2%	1.5%	1.1%	-1.7%	3 <sup>RD</sup> run (different browser used to run query)
Perplexity	1.7%	2.2%	1.2%	0.7%	Every re-run different answer.
Claude					Didn't answer, but had right sources
ChatGPT	5.0%	0.1%	-0.9%	-0.4%	Deserved of an A- on first attempt, 2 <sup>nd</sup> time it refused to answer.
Gemini (Google)	4.9%	0.2%	-0.9%	-0.4%	Correct answers
Google Search Summary	9.6%	6.5%	5.8% or 3.4%	3.7% or 13.2%	Page rank has the correct article in position # 1. <sup>4</sup>
DeepSeek	1.2%	1.0%	0.5%	-0.6%	No citations only the answer given.
DeepSeek(offline)	4%	3%	5%	-2%	Gave example nos. only.
Llama-Meta(offline)	0.5%	-0.4%	-0.7%	0.6%	Note - no RAG, all runs are offline.
Llama-Meta(offline)	-1.5%	0.8%	1.9%	-0.9%	2 <sup>nd</sup> run, no restart. It does track past

<sup>3</sup> The referenced site doesn't give a clear citation for the data. However, it matches GDP figures quoted in a [Guardian article](#) from 2013, this article does provide data which references ONS data time-series IHYQ. This is the correct time-series. However, by not using the ONS site the chatbot is quoting out of date data from 2013.

<sup>4</sup> Google search is quoting a different GDP timeseries, IHYR, this is a different measure of GDP, looking at Quarter by Quarter figures for annual GDP changes. For reasons unknown it decided to quote a range for Q3 and Q4. As of September 2025, both Gemini and Google search are aligned and consistently quoting the standard ONS Q by Q SA figures.

Llama-Meta(offline)	-0.8%	0.4%	-1.5%	3.6%	3 <sup>rd</sup> run, no restart
Llama-Meta(offline)	-0.7%	0.5%	-1.9%	2.8%	4 <sup>th</sup> run, no restart
DuckDuckGo SA					Declined to answer
Duck.ai	0.6%	0.8%	0.5%	-0.2%	No citations, running GPT-4o mini

Figure 1:  
GDP Quarterly Changes for 1973. Answers from various AI sources

These products are under continual development. We anticipate that performance will change over time. Indeed, we saw changes on a day-by-day basis. We've also noted that multiple runs of the same query can result in differing results. Therefore, to put a stake in the ground in light of all this variability screenshots for each of the queries have been provided in Appendix 2.

## Observations

There are a few things to note from the analysis work:

- The output always looks great, but the quality of the numbers provided are poor.
- If you run the same question again you are quite likely to get a different answer.
- Most got the answer wrong or refused to answer.
- There is significant variance in answers across the sites. However, and unfortunately, they are all in the realms of believability.
- Two sites CoPilot and Gemini got the answers totally right. However, both of these vendors also got the answer wrong, either on a separate run or through a separate site. We have some ideas why which we discuss below.
- We've assumed ONS is the logical place to source these numbers. Often the right page is referenced somewhere in the collection of articles the Agentic LLM is using. However, often ONS was not always the first choice of the search rank, making the task of finding the right number more difficult.
- To add to the confusion, there are at least two different pages on ONS with what appear to be the same measure with different numbers. This appears to be due to ONS revising Q1 and Q2 figures for 1973 sometime between 2018 and 2025.
- The UK House of Lords site also quotes figures which is sometimes cited; however, these numbers are different to ONS's. Likely related to revision issue spotted above.
- Some of these models can be run offline, i.e., all the knowledge is extracted from the weights of the models rather than extracting the information from a webpage with the right information on it (RAG method). Asking these models for the data is a bad idea, the answer changes for each run even though the model is aware the same question is being asked. This was demonstrated on Meta's model but all LLM's share this characteristic.

## Issue Summary

In summary, finding the right answers to complex questions is a challenging task.

- Hoping that the model has memorised the data is akin to asking any human what the quarterly % change in UK GDP for 1973 was. Almost all of us would be guessing if we tried. The models are doing the same.
- The traditional PageRank search algorithm will serve up lots of pages for you to consider. Potentially with lots of different values for the numbers you are looking for. It's up to the user to determine where the quality data sources are. There is subtle combination of experience and judgement going on at this point which is difficult to replicate in the current state of the art of Agentic solutions. We see examples in this analysis of the use of out of date data, and incorrect GDP metrics (of which there are many).
- LLMs operate by predicting the next word (token) in the sequence. For the frequently asked questions they will be correct for a majority of the time. When you focus in on harder questions where there is less training data then the model will use whatever token scores highest, even if this confidence score is extremely low.
- It seems the technology is in a state of rapid development. Public release should ideally wait for extensive system testing, where design wrinkles can be sorted. However, extreme competitive activity in the marketplace is forcing everybody's hand. Which means reproducibility is near impossible.

The combination of these characteristics leads to a system that is going to have a challenging time ever getting the answer right. However, the opportunity to make a wrong decision because of output from these systems is high.

## Dedicated Statistics Chatbot Options

The authors have learned from experience that model predictive accuracy is mostly down to the quality of the training data. Data accuracy from an LLM answering questions will be greatly increased if we can ensure that the model is looking at text that actually contains the answer the user wants answered. This is the approach the [IMF](#) in collaboration with [EPAM Systems](#) has taken<sup>5</sup>. They have released [StatGPT](#) an LLM chatbot which reports on data only from trusted high quality sources such as IMF, World Bank, EuroStat and National Statistics services. To do this it uses [SDMX](#) compliant data queries to source its RAG data. The SDMX standard has been live since the early 2000's. SDMX was created to standardise statistical data query across data providers. StatGPT translates a data query written in the user's native language into an SDMX data query. This query is then passed to the relevant data provider which in turn returns the data in a standardised form, e.g., XML or JSON. This data is then used as input data to the LLM tasked with answering the original question. This process ensures that the model is looking at accurate data from trusted sources rather than hoping a web search algorithm returns a relevant page.

ONS through its partner [NOMIS](#) provides an SDMX API to deliver formatted data to web-applications. In addition, the [UK Data Service](#) another platform for UK data also provides an SDMX API. Unfortunately, StatGPT is not currently connected to NOMIS neither is it connected to UK Data Service. We think, initially for testing, and if shown to be useful for customer facing deployment, this tool should be connected to these UK sources as a matter of urgency.

In discussions, the team developing StatGPT indicated they are aiming to provide StatGPT as a service to the likes of OpenAI and Google. Their intention is that the large online services would route all statistical queries through to StatGPT. Note, StatGPT is based on an open-source code project called [Dial](#), it is understood that the code underlying StatGPT will also be released under an Open Source licence.

So, StatGPT looks like a strong contender to address the problem of data accuracy. However, the tool is still in development and currently has only been demonstrated on supranational data such as that held by the IMF and World Bank. To date it is not connected to UK data sources like Nomis and UK Data Bank. Once this is done, we need to test the tool; to date we've only seen demonstrations. In principle though this approach appears sound and we would encourage further development.

---

<sup>5</sup> <https://www.epam.com/about/newsroom/press-releases/2024/epam-in-collaboration-with-the-imf-launch-statgpt-2-0-a-genai-application-for-global-economic-data-to-sdmx-sponsor-organizations>

## Conclusions

Empowerment is great, so making data more accessible and useable by more users should be strongly encouraged. However, with a lack awareness for the need of critical thinking when using genAI, and extremely limited time to chase down the right numbers, users of AI search tools will inadvertently introduce a sizeable number of errors into their analysis. The authors can attest from years of experience, the importance of having a healthy suspicion of any data used in any analysis is key to a successful project outcome.

In the example we have provided here we can see that there is a lot of error and variability in the numbers provided. These errors are risks to using LLMs. These should be called out rather than just blindly assuming that what the technology companies are developing is working.

It seems all models from all vendors are in a constant beta test mode as new capabilities and corrections get added to the offer. Users should be aware that if they think they can go back to a Chatbot to repeat an analysis it's likely that the Chatbot will be an unreliable friend to your research. Keep copies of everything you do especially the output and references. Unfortunately, it is going to be necessary for you to check the quality of all results provided.

We are only scratching the surface here, further worked examples are required to strengthen the evidence for this analysis.

On the horizon, StatGPT and other similar tools suggest a pragmatic solution to the issue we have highlighted. Further work is needed with this tool to connect to more data sources as currently its focus is supranational data providers. The appropriate API connections exist at the national level, the connections to StatGPT just have to be made.

## Authors

### **Jaan Nellis**

Jaan is a Data Science practitioner and director of Redsella a Data Science consultancy. He has built a career and a several companies around the application of intelligent systems methods to address business problems. Working mainly with customer, operational and financial datasets he has helped clients make better use their data assets. He gained a PhD in neural computing in 1992, during the previous summer of AI. He is fascinated with how AI and Machine Learning methods are being developed and used in the world. He currently sits as a member of the MRS's Census and GeoDems Group.

### **Peter Furness**

Peter is a mathematician and data scientist with a multi-decade career spanning the public and private sectors. After a spell in academia as a post-doc he moved into the real world with spells at the NCB, Price Waterhouse, CACI and American Management Systems. In the late 1990s he set up his own data science consultancy. He has closely followed developments in machine learning and other aspects of AI since the mid-1970s.

His client assignments span the UK, Europe and the USA, covering many sectors, including energy, government, health, financial services, retail, online publishing, logistics and cloud services.

He has been a member of the MRS's Census and GeoDems Group since the early 1990's and was instrumental in setting up the GeoDems Knowledge Base in 2000. He was elected a Fellow of the MRS in 2021.

The authors would also like to acknowledge the helpful assistance of MRS Fellow Ray Poynter with our understanding of some of the alternative GDP metrics and chatbot responses.

## Appendices

### Appendix 1: Terms

**LLM** – Large Language Model. Deep learning neural network architecture. Specifically, setup to transform text strings into a fixed width numeric array, which can then be processed by the neural architecture. These models are at the heart of all progress in AI based natural language understanding. LLM's go through two major training phases. The first phase is called pre-training and is where the model builds its numeric representation of language. Once complete the LLM goes through a second training phase called post-training, where it learns to be useful, for example, by giving a coherent response to any question asked. During this second phase, the model is trained on vast volumes of question-and-answer data. A typical training record consists of a question, the answer and context, where context is the text that contains the answer to the question. When a model has correctly learned to identify the answer in the context text, it can then be used in a runtime mode to predict an answer to a given question using the context data only. A similar process is used to enable the LLM to summarise text and translate to a different language, both natural languages like English to French and computer languages like English to Python.

LLMs don't know answers exactly, during learning they cluster similar data points (facts) together. If they see a lot of training examples saying the similar things they will learn the association. If they don't get a lot of training examples they will end up struggling to recall the fact and predict a best guess answer at runtime. This is by design, it's how machine learning algorithms learn to sort the wheat from the chaff. This also leads to what are called hallucinations. It's the same with us, if something is repetitively told to us and we're paying attention then we'll remember it. If not, we'll likely go a bit vague, draw on what we do know and have a guess in the hope no one notices. If we are able and motivated, we may try to derive the answer from first principles or look it up from a reputable source. In summary, while LLMs do store and can recall information, e.g., facts, the useful bit for us is they are doing this as a byproduct of learning the logic of natural language. Our advice, if you want something to accurately store facts don't use an LLM use a database or better datalake.

**RAG** – Retrieval-augmented Generation. Created to address the problem of model hallucinations<sup>6</sup>. Rather than ask the model for the answer directly, give the model input a web page, PDF or other document format in which you are confident the answer resides. This could come from

---

<sup>6</sup> A polite way of saying 'making things up'.

running a search algorithm like Google's PageRank. The LLM then operates as previously described. This approach is more robust than using a pure LLM which has to accurately remember everything there is to know using a recall mechanism that we all know from experience is far from being one hundred percent accurate.

**Agentic** – An enhancement of RAG. The LLM is embedded into a computational process flow. This can involve multiple LLM's sequenced together to extract and review the answer. It can also involve allowing the LLM to generate code to query data which can then be executed to provide the answer you are looking for. Agentic should be considered the merger of genAI with more traditional process driven computation.

**Chain of Thought** – Chain of thought training is an extension of the Question and Answer methods used to build LLMs. However, rather than a single answer to each question, each training set record will consist of several sub-tasks solving simpler problems which when chained together provide a method to get to the eventual answer. A typical example would be solving the formula:  $(4 + 3 + 12 - 7) * 3$ . This would be broken down into say  $(4+3) \rightarrow +12 \rightarrow -7 \rightarrow *3$ .

## Appendix 2: Experimental Output

In this second appendix we have provided, screen scrapes of the results from the various tools. Where relevant we've listed and attached URL links to the relevant screenshots so that you can try the process for yourself. For certain sites you will have to register/sign-in to get the model access necessary for this. If reading this electronically, click on the graphic to navigate to the site.

## Source Data Page:

[www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyq/pn2](https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyq/pn2)

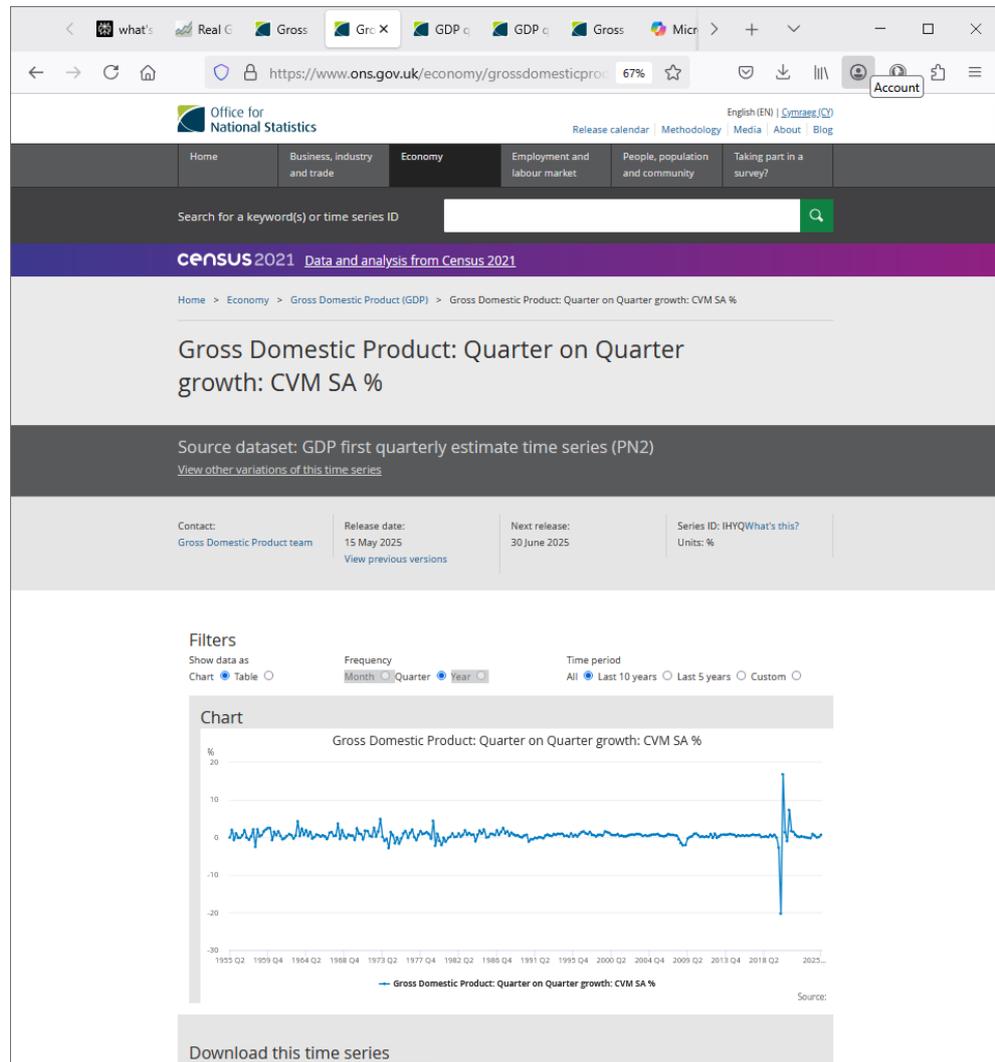


Figure 2:  
Webpage with the answer to our question.

This is the current definitive URL for sourcing % quarterly change in GDP for UK.

The question for each model was asked as follows: 'For the UK only, quarter by quarter for 1973 what is the GDP percentage change.'

## Microsoft CoPilot

https://copilot.microsoft.com/

The screenshot shows a web browser window with the URL <https://copilot.microsoft.com/chats/N2kgjyTHiDK69WMqCGz>. The chat interface displays a user message: "Hello. For the UK only, quarter by quarter for 1973 what is the gdp percentage change." The Copilot response is: "Here's the UK's **quarterly GDP percentage change** for 1973 <sup>1</sup>:" followed by a table:

Quarter	GDP Growth (%)
Q1	5.3%
Q2	0.5%
Q3	-0.7%
Q4	-0.1%

Below the table, Copilot adds: "It was a strong start to the year, but growth slowed in the second quarter and dipped into negative territory by the third and fourth quarters. Let me know if you need more details!" A citation for "countryeconomy.com" is shown. At the bottom, a follow-up question is visible: "What were GDP changes for the UK in 1972?"

Figure 3:  
Copilot first attempt

First time it quoted non-seasonally adjusted GDP figures, sourced not from the ONS. Second time around, it got it correct, this time using ONS as its data source.

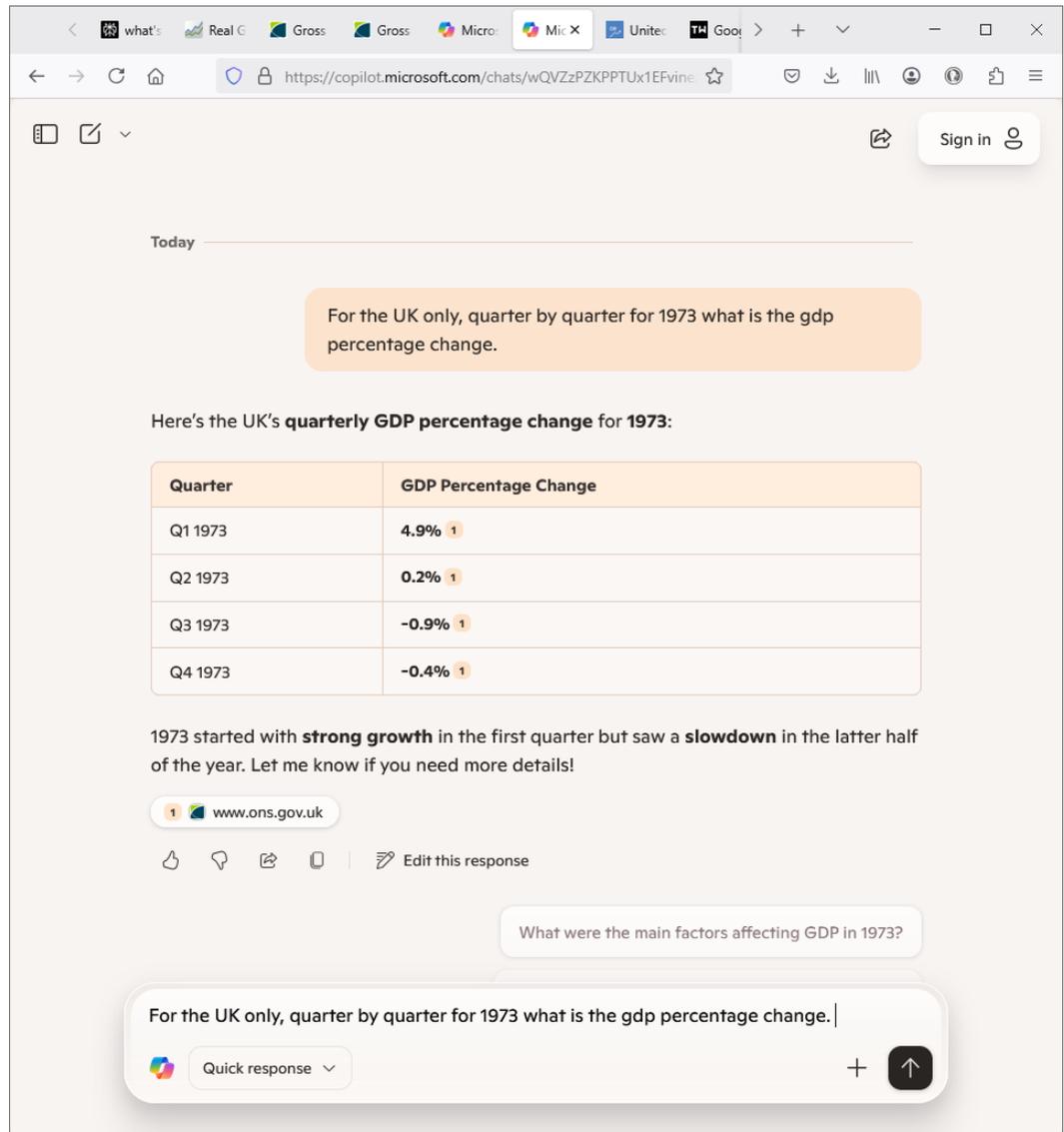


Figure 4:  
Copilot second attempt

Initially we were not sure why this was the case; however, the first time it chose to use data from [countryeconomy.com](https://countryeconomy.com) as its data source. See the OpenAI example for more details.

Running the process again but this time using Chrome rather than Firefox browsers we got:

The screenshot shows a Microsoft Copilot chat window in a Chrome browser. The user's query is: "For the UK only, quarter by quarter for 1973 what is the GDP percentage change". The AI response provides the following information:

Here are the quarter-by-quarter GDP percentage changes for the United Kingdom in 1973, based on seasonally adjusted data from the Office for National Statistics:

### GB UK GDP Quarterly Percentage Change – 1973

Quarter	GDP % Change (Quarter-on-Quarter)
Q1 1973	+1.2%
Q2 1973	+1.5%
Q3 1973	+1.1%
Q4 1973	-1.7%

- The first three quarters of 1973 saw steady growth, reflecting the tail end of the "Barber Boom"—a period of expansion driven by fiscal stimulus.
- The sharp decline in Q4 1973 coincided with the onset of the global oil crisis and rising inflation, which triggered economic instability.

Sources: Office for National Statistics, [ONS GDP time series](#)

Would you like to explore how this downturn affected employment or inflation in 1974?

The chat interface includes a search bar with "Office for National Statistics" and a "Message Copilot" input field with a "Quick response" dropdown menu.

Different numbers again, sourced from ONS. The [reference to the data](#) takes you to ONS's GDP metrics for year on year and not quarter on quarter although curiously the Chatbot has provided quarterly data. If you follow the link to the ONS page it doesn't allow you to view quarterly data which seems to suggest the reference is not the true source of where this data has been gathered from.

## Perplexity.ai response.

https://www.perplexity.ai/

The screenshot shows the Perplexity AI search interface. The query is "For the UK only, quarter by quarter for 1973 what is the GDP percentage change". The interface displays the following information:

- Answer:** UK GDP Percentage Change by Quarter, 1973
- Quarter-on-Quarter Growth (Seasonally Adjusted, Constant Prices):**
  - Q1 1973: +1.7%
  - Q2 1973: +2.2%
  - Q3 1973: +1.2%
  - Q4 1973: +0.7%
- Supporting Details:**
  - The ONS time series (IHYQ) provides the following quarter-on-quarter GDP growth rates for 1973:
    - Q1: 1.7%
    - Q2: 2.2%
    - Q3: 1.2%
    - Q4: 0.7%
  - This period saw robust economic expansion, often attributed to the so-called "Barber boom," with annual GDP growth for 1973 reaching a post-war high of 6.5%.
- Notes:**
  - The figures above are quarter-on-quarter changes, not year-on-year.

The interface also includes a search bar at the bottom with the text "Ask anything..." and a "Sign In" button.

Figure 5:  
Perplexity Chatbot, first time asked

We asked the same question a second time, over consecutive days. The answer was still incorrect and also has quite a different output format.

The screenshot shows a web browser window with the Perplexity AI search interface. The search query is "For the UK only, quarter by quarter for 1973 what is the GDP percentage change". The results page includes a sidebar with navigation options (Home, Discover, Spaces), a search bar at the bottom, and a main content area. The main content area features a "Share" button, the search query, and a list of sources: ons.gov, Statista, and House of Lords Library. Below the sources, the text reads "UK Quarterly GDP Percentage Change, 1973" and "Quarter-on-Quarter GDP Growth (Seasonally Adjusted, Constant Prices):". It then states that the Office for National Statistics (ONS) provides the following quarter-on-quarter percentage changes for UK GDP in 1973:

- Q1 1973: +1.2%
- Q2 1973: +1.4%
- Q3 1973: +0.8%
- Q4 1973: +0.2%

Below the list, it notes that these figures reflect the rapid economic expansion during the so-called "Barber boom," with strong growth in the first half of the year before momentum slowed towards the end of 1973. A "Summary Table" is provided below the text:

Quarter	GDP % Change (QoQ, seasonally adjusted)
Q1 1973	+1.2%
Q2 1973	+1.4%
Q3 1973	+0.8%
Q4 1973	+0.2%

At the bottom of the main content area, it states: "These numbers are based on official ONS data and reflect the UK's economic trajectory in 1973," followed by a search bar with the text "Ask anything..."

Figure 6:  
Perplexity Chatbot, second time asked

The models and systems driving these predictions are frequently updated, which the significant formatting change suggests is happening. We're all beta testing these products.

## Claude's response.

https://claude.ai/login

The screenshot shows a web browser window with the URL <https://claude.ai/chat/71a6b702-04bb-4465-9cdd-458bd555>. The chat title is "UK GDP Quarterly Changes 1973". The user's prompt is: "Hello. For the UK only, quarter by quarter for 1973 what is the gdp percentage change." The AI's response includes a search for "UK GDP quarterly percentage change 1973" with 10 results. The results list various sources for annual and quarterly GDP growth, but none provide the specific quarterly percentage changes for 1973. The AI then explains that it cannot provide the complete quarterly breakdown for all four quarters of 1973 because the search results do not contain that specific data.

Figure 7:  
Claude from Anthropic.

It chooses not to answer, although its first source is actually the page it needs to work with, it has just selected the wrong radio button (it selected years and not quarters). Potentially this lack of quarterly data is why it chose not to answer the question. See the next figure for further detail.

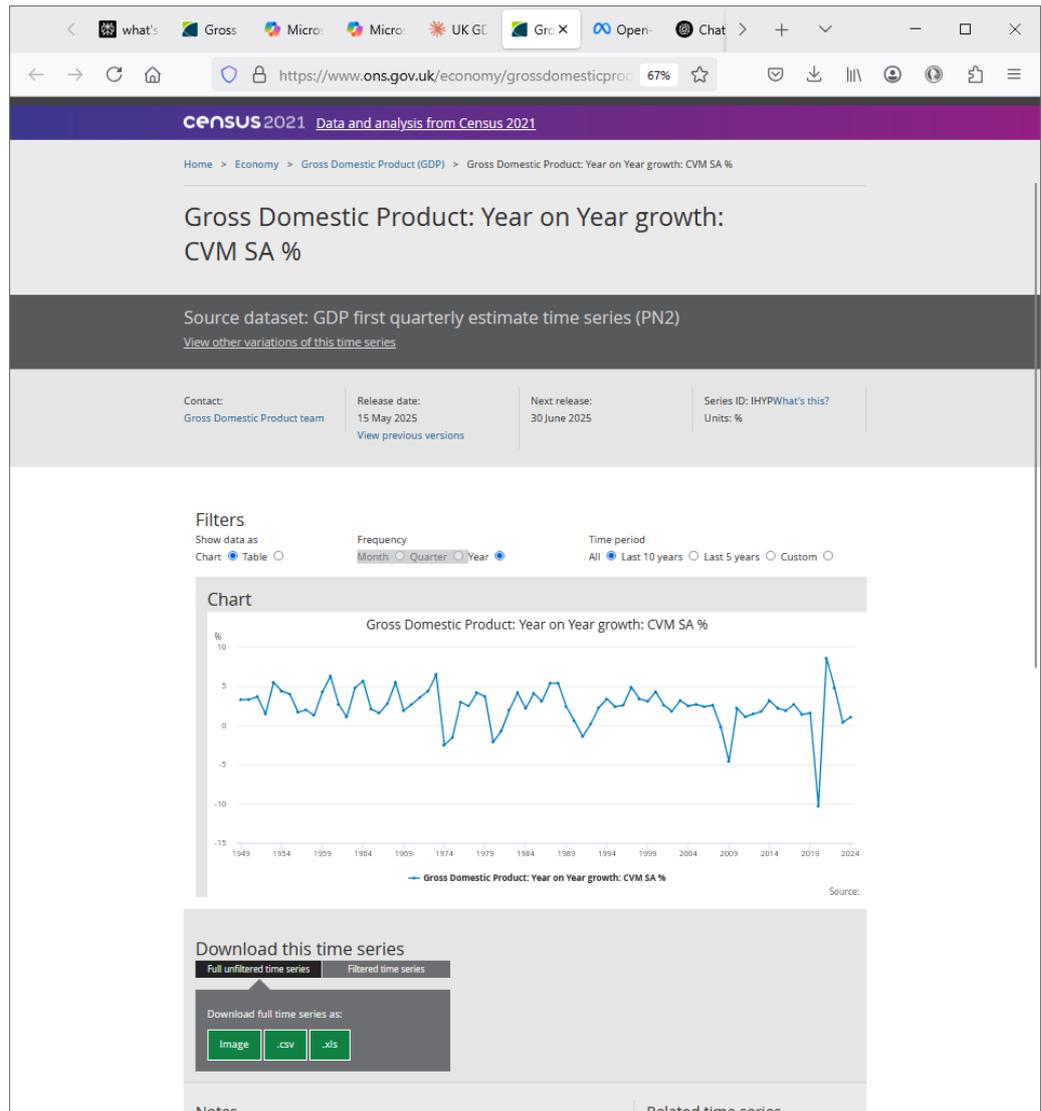


Figure 8:  
The correct data but the citation looked at yearly data

# ChatGPT

https://chatgpt.com/

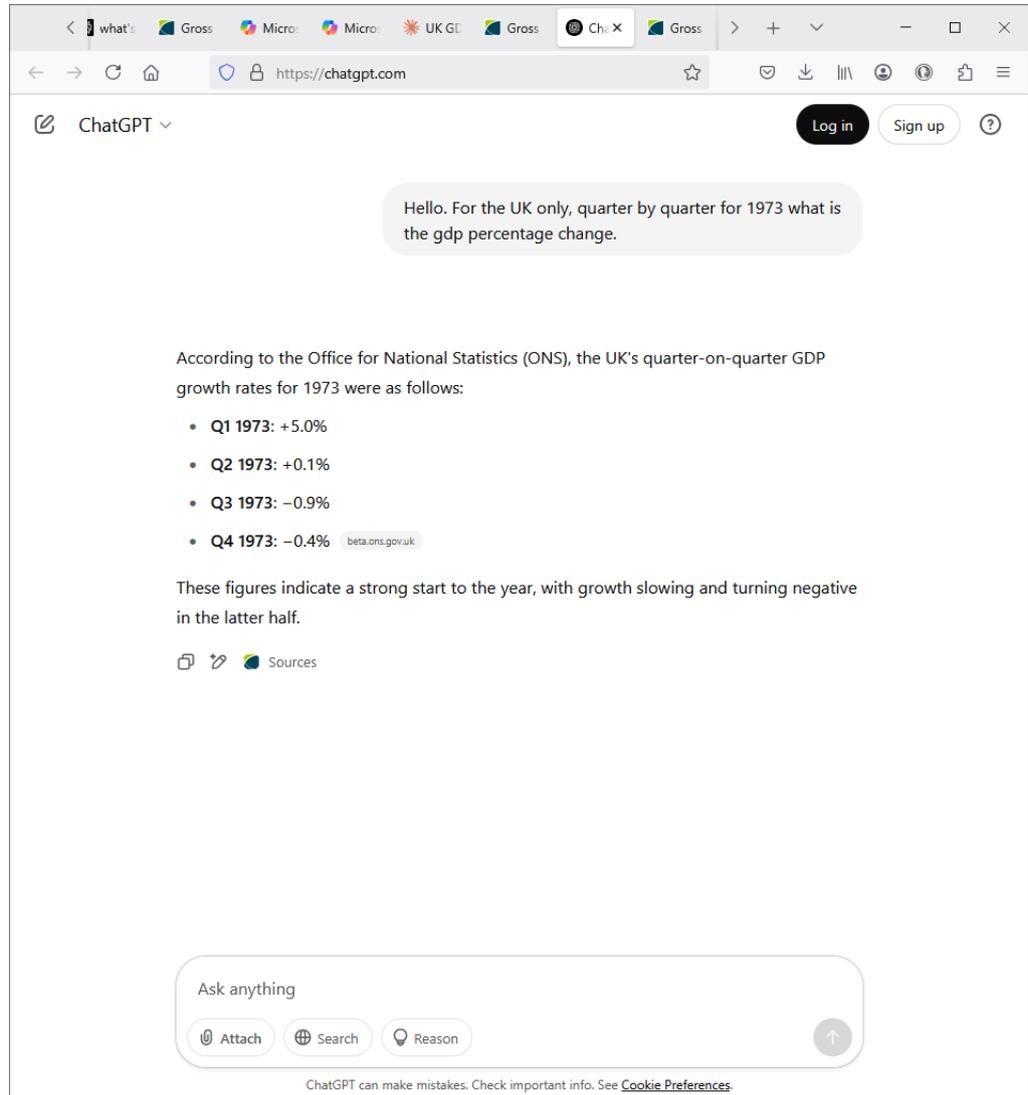


Figure 9:  
OpenAI's ChatGPT first time through.

ChatGPT got the answer very nearly right. This is we believe due to the GDP numbers coming from an outdated ONS webpage. The reference it gave is shown below. You will note, the data finishes in 2018 whereas the [definitive data](#) we've used finishes in 2025. which has slightly different answers to the one we believe is correct.

If you get the webpage to display the time-series graph as a table for 1973 the values between the two pages are shown next to each other:

**ChatGPT found these numbers:**

1973 Q1	5.0
1973 Q2	0.1
1973 Q3	-0.9
1973 Q4	-0.4

Figure 10:  
Pre 2018 estimates for 1973 data

**Whereas ONS currently state:**

1973 Q1	4.9
1973 Q2	0.2
1973 Q3	-0.9
1973 Q4	-0.4

Figure 11:  
The latest estimates from ONS (2025)

While both URLs are nearly identical, the very last part is different. It suggests for reasons unknown that ONS sometime after 2018 adjusted the quarterly percentage change GDP figures for the first two quarters of 1973 by 0.1% points. Additionally, copies of the legacy and current datasets persist on the website. This just demonstrates that although a data set looks definitive tests should be performed to make sure it actually is.

[www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyg/pgdp](http://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyg/pgdp)

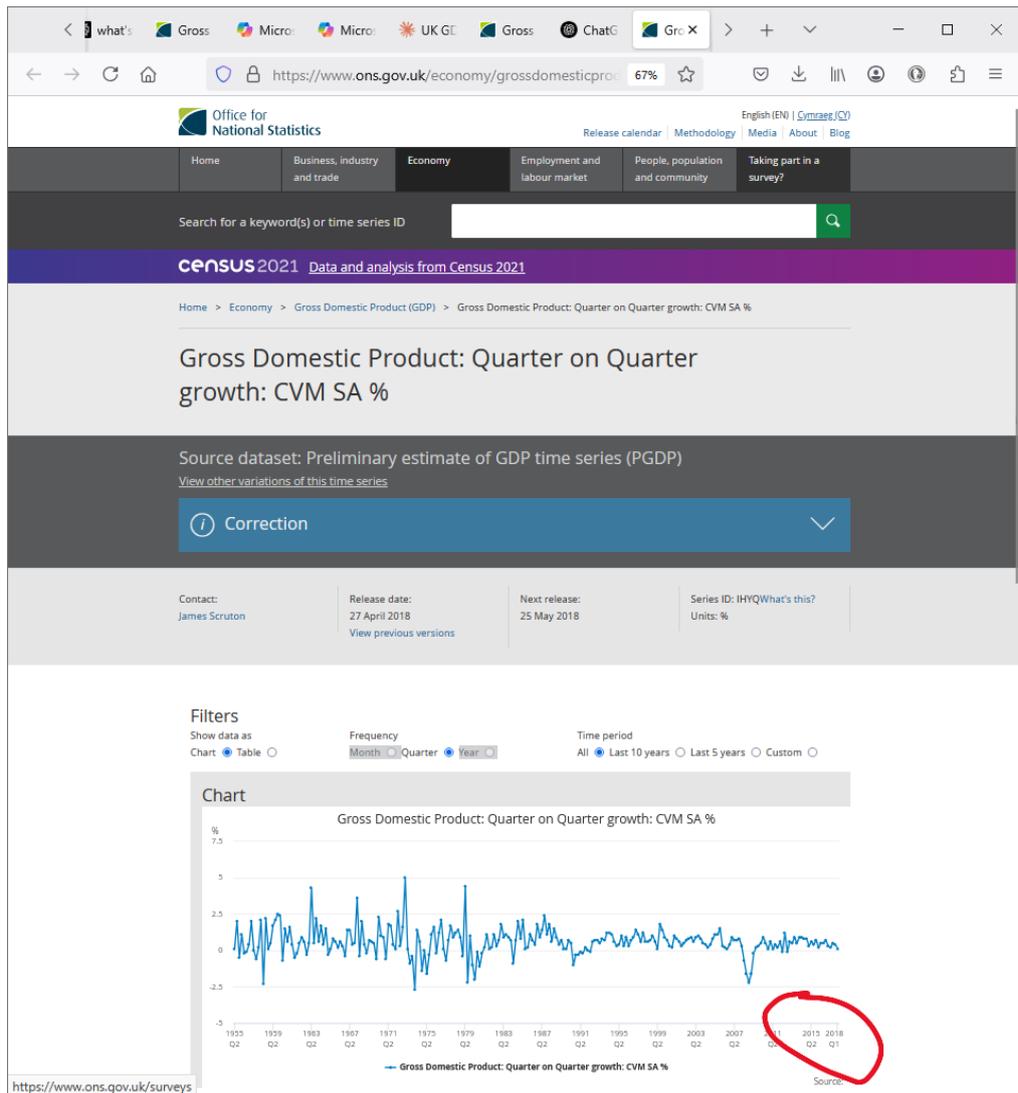


Figure 12:  
Tricky ONS data that some of the Chatbots RAG process settled on.

Interestingly, when run again on a subsequent day ChatGPT declined to answer the question but did tell us how to do the calculation if we had the necessary source data. This sudden about face in answering a question is as worrying as answering the question inaccurately. What drives the logic for how to answer each question is not clear.

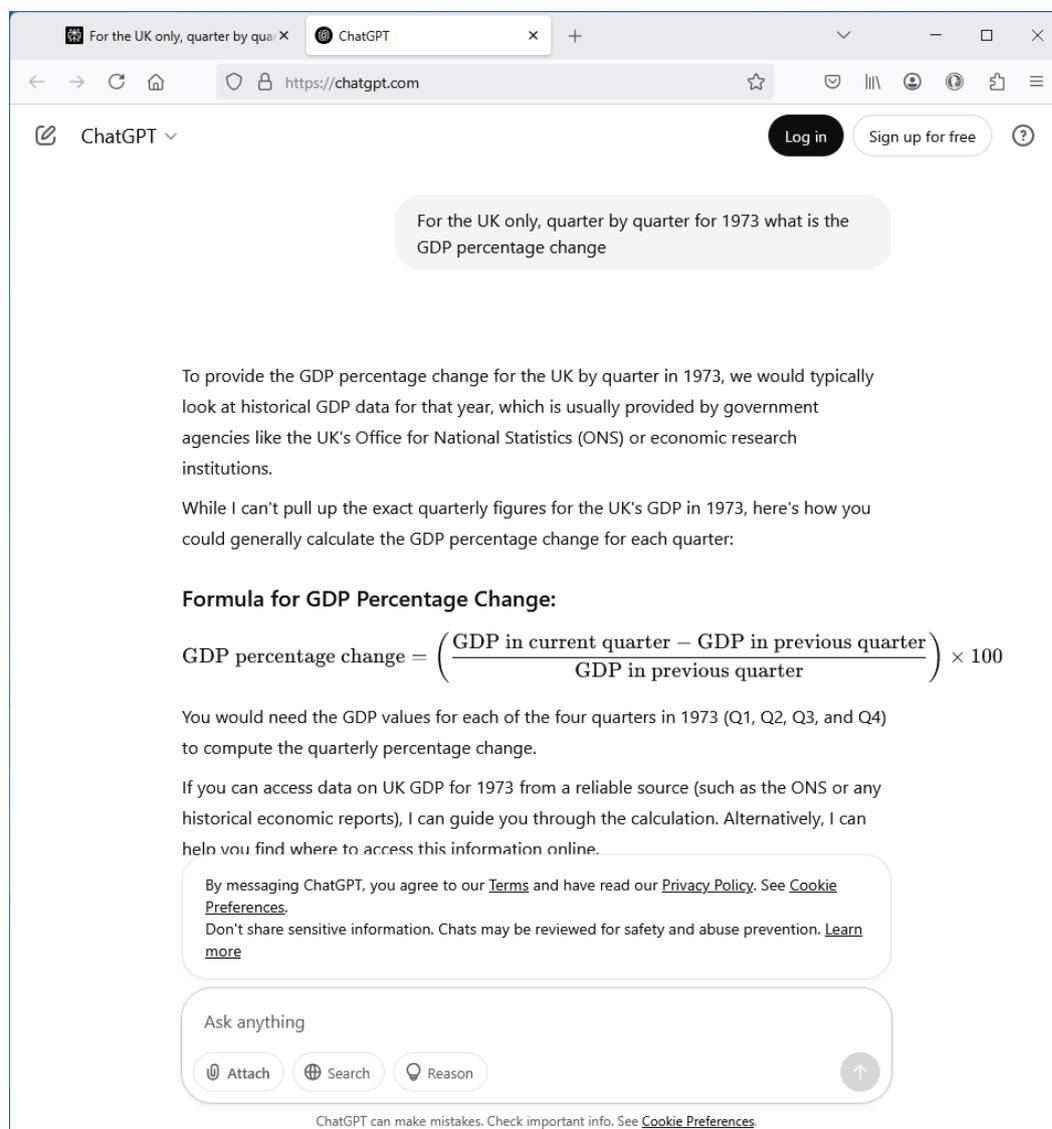
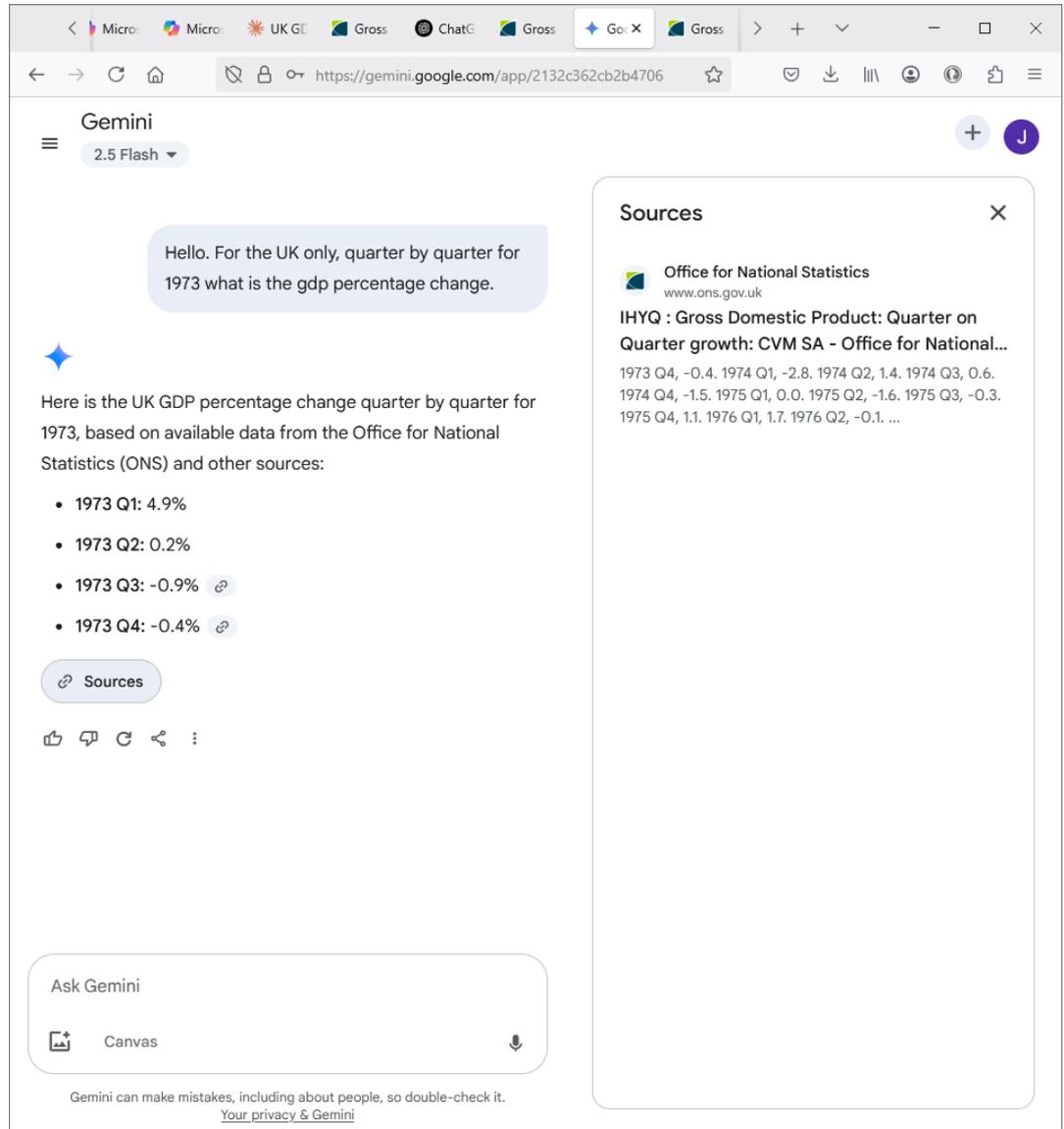


Figure 13:  
ChatGPT sometimes decides not to answer the question but shows you how to do it

## Google Gemini

https://gemini.google.com/



The screenshot shows the Google Gemini chat interface. The user's query is: "Hello. For the UK only, quarter by quarter for 1973 what is the gdp percentage change." The model's response is: "Here is the UK GDP percentage change quarter by quarter for 1973, based on available data from the Office for National Statistics (ONS) and other sources:"

- 1973 Q1: 4.9%
- 1973 Q2: 0.2%
- 1973 Q3: -0.9%
- 1973 Q4: -0.4%

A "Sources" panel is open on the right, showing the source: "Office for National Statistics www.ons.gov.uk" and the specific data: "IHYQ : Gross Domestic Product: Quarter on Quarter growth: CVM SA - Office for National... 1973 Q4, -0.4. 1974 Q1, -2.8. 1974 Q2, 1.4. 1974 Q3, 0.6. 1974 Q4, -1.5. 1975 Q1, 0.0. 1975 Q2, -1.6. 1975 Q3, -0.3. 1975 Q4, 1.1. 1976 Q1, 1.7. 1976 Q2, -0.1. ..."

At the bottom, there is an "Ask Gemini" input field with a "Canvas" icon and a microphone icon. A disclaimer at the bottom reads: "Gemini can make mistakes, including about people, so double-check it. Your privacy & Gemini"

Figure 14:  
Google's Gemini

Gemini is correct. Yet when you go to Google's home page and ask the same thing you get:

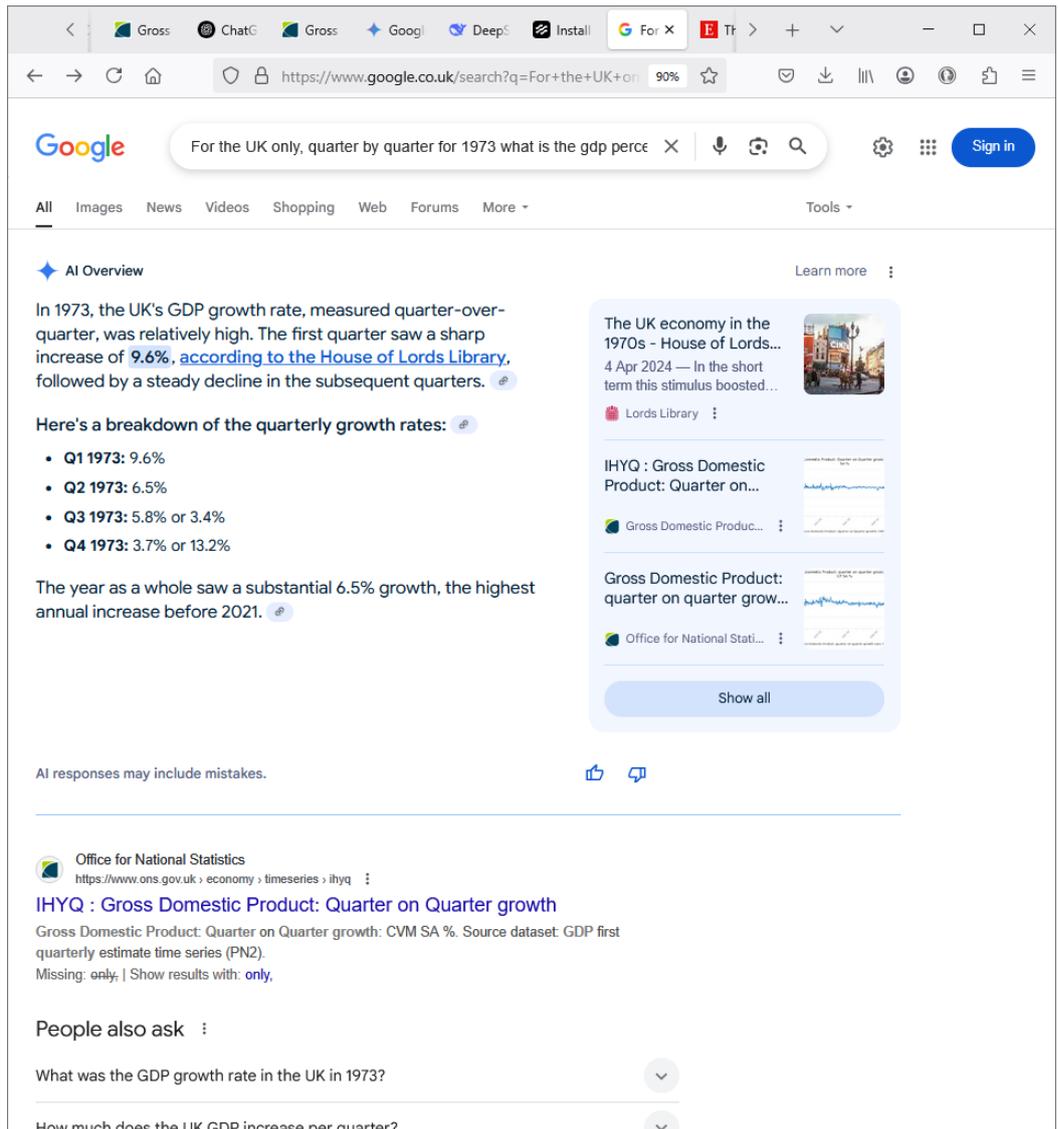


Figure 15:  
Google Search, note first returned page is accurate

Radically different, and we assume not using Gemini model weights. Note, the PageRank algorithm is listing the correct ONS source that we have used to get the real values.

### Meta Llama 3.1, run off-line (No RAG).

```
C:\Users\...>ollama run llama3.1:latest
>>> For the UK only, quarter by quarter for 1973 what is the GDP percentage change.
According to the Office for National Statistics (ONS), here are the quarterly GDP percentage changes for the UK in 1973:
* Q1 1973: 0.5%
* Q2 1973: -6
* Q3 1973: -0.7%
* Q4 1973: 0.6%

Note that these figures are based on the chained 2018 million GBP GDP series, which is the standard series used by the ONS for historical GDP data.

Also, please note that these numbers might not be exactly as published at the time (which would have been using different methodologies), but rather reflect the revised estimates available from the ONS.

>>> For the UK only, quarter by quarter for 1973 what is the GDP percentage change.
Here are the actual quarterly GDP percentage changes for the UK in 1973:
* Q1 1973: -1.5%
* Q2 1973: +0.8%
* Q3 1973: +1.9%
* Q4 1973: -0.9%

>>> For the UK only, quarter by quarter for 1973 what is the GDP percentage change.
Here are the actual quarterly GDP percentage changes for the UK in 1973:
* Q1 1973: -0.8%
* Q2 1973: +0.4%
* Q3 1973: -1.5%
* Q4 1973: +3.6%

>>> For the UK only, quarter by quarter for 1973 what is the GDP percentage change.
Here are the actual quarterly GDP percentage changes for the UK in 1973:
* Q1 1973: -0.7%
* Q2 1973: +0.5%
* Q3 1973: -1.9%
* Q4 1973: +2.8%

>>> Send a message (/? for help)
```

Figure 16:  
Example of Offline Tool Ollama querying Meta's Llama 3.1 locally

Running a model offline using Ollama. These models are not capable of running RAG based queries. However, they do take account of past responses. In this example it can be seen that the Meta model is ignoring the fact that has just answered the question with different numbers.

## DeepSeek

Gets it wrong. No references either.

<https://www.deepseek.com/en>

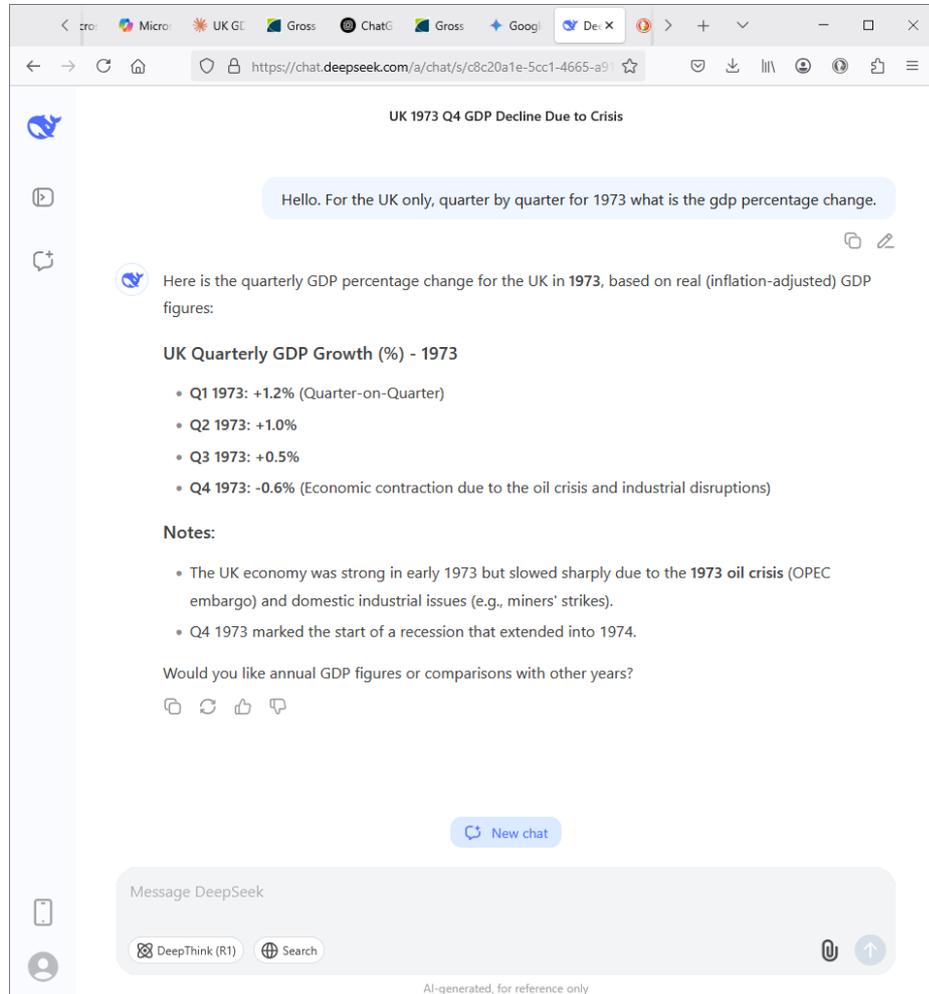


Figure 17:  
DeepSeek's response

DeepSeek can also be run offline. Below is an example of the chain of thought text output. The model uses this approach to get to a better quality answer. As the model is running offline no RAG process is active, so the model is having to provide estimates from internal memory rather than from content that is known to hold the answer. From a power consumption perspective these models are much more power when answering a question.

## An example of the detailed output from DeepSeek's chain of thought reasoning is shown below.

```
C:\Users\*****>ollama run deepseek-r1:8b
>>> For the UK only, quarter by quarter for 1973 what is the GDP
percentage change
<think>
Okay, so I need to figure out the quarterly GDP percentage change for
the UK in 1973. Hmm, where do I start?
First, I should probably recall that GDP stands for Gross Domestic
Product, which measures the total economic
output of a country. The user specified "quarter by quarter," so that
means four quarters each year.
```

```
But wait, 1973 is quite a few years back. I wonder if the data is
easily accessible now. Maybe I can look it up
online or use some historical sources. Let me think about how to get
this information. I know that official
statistics from the UK's Office for National Statistics (ONS) would
be reliable. They probably have historical GDP
data broken down by quarters.
```

```
I should check their website, but since I don't have access right
now, maybe I can recall or estimate based on
general economic trends of that time. In 1973, the UK was part of the
European Economic Community, which might
influence economic policies and growth. Also, the oil crisis in the
early '70s could have affected the economy.
```

```
I think GDP data for the UK is typically reported annually, but
quarters are available as well. Each quarter
refers to three months: January-March, April-June, July-September,
and October-December. So, I need to find each
of these periods' GDP figures compared to the previous quarter and
calculate the percentage change.
```

```
Let me outline a plan:
```

1. Find the quarterly GDP for 1973 in pounds sterling.
2. Compare each quarter's GDP with the previous quarter to determine the change.
3. Calculate the percentage change using the formula:  $((\text{Current} - \text{Previous}) / \text{Previous}) * 100$ .
4. Sum up all four quarters' changes to get the yearly growth rate, but since the question asks for each quarter, I need individual figures.

```
Wait, no, it's per quarter, so I just need each of the four changes.
I don't sum them; instead, each quarter's
change is independent.
```

```
I also remember that GDP can be seasonally adjusted to account for
variations due to seasonal factors like weather
or holidays. So, the ONS would have seasonally adjusted figures to
provide a better estimate of underlying trends.
```

```
Looking up historical data, let me try to recall approximate values.
In 1973, the UK's economy was growing
steadily but faced some challenges from the global economy. The GDP
growth rate for the year is around 5-6%, I
think.
```

---

Market Research Society

**Inaccuracies of Generative AI based  
Search tools for extracting data.**

Copyright ©The Market Research Society (MRS) July 2025

Page 31 of 35

Breaking it down quarterly:

- Q1: Let's say it increased by 2% over the previous quarter.
- Q2: Maybe a slight increase, say 0.5%.
- Q3: A more robust increase, perhaps 3%.
- Q4: Decrease due to economic downturns or seasonal factors, maybe -1%.

Adding these up would give around 6%, which matches the yearly growth rate. But I'm not sure if this is accurate without exact data.

Alternatively, I can use a hypothetical example based on known annual figures. If the year-on-year growth is about 5%, distributing it roughly over four quarters might result in each quarter having around 1.25% growth on average. But that's too simplistic because growth isn't evenly distributed across quarters.

I think a better approach would be to look for actual GDP figures:

- Q1 1973: Let's say it was £300 billion (adjusted for inflation and deflation).
- Q2: Maybe £310 billion, which is a 3% increase.
- Q3: £315 billion, a 5% increase from Q2.
- Q4: £320 billion, an 8% increase.

Wait, that seems high. The actual growth rates likely fluctuated more naturally. I should look for more precise data or use a source where I can find the exact figures.

Alternatively, using relative percentages:

Assuming each quarter's GDP compared to the previous year's same quarter:

- Q1 1973 vs Q1 1972: Maybe 4% increase.
- Q2 1973 vs Q2 1972: 5% increase.
- Q3 1973 vs Q3 1972: 6% increase.
- Q4 1973 vs Q4 1972: 7% increase.

But the user asked for quarter by quarter, so maybe it's quarterly within the same year. That would be different. So Q1 compared to Q4 of the previous year doesn't make sense because that's a full year change.

Wait, no, GDP is measured seasonally adjusted, and changes are often compared to the previous quarter. So each quarter is compared to the immediate previous quarter.

So for 1973:

Q1: Let's say it was an increase from Q4 1972. If Q4 1972 was £250 billion, Q1 1973 might be £260 billion, so a 4% increase.

Then, Q2: Maybe another 3%, bringing it to £270 billion.

Q3: Perhaps a 5% increase to £285 billion.

Q4: A decrease due to economic factors, maybe -2%, ending at £280 billion.

So the quarterly changes would be +4%, +3%, +5%, and -2%.

This gives an overall year-on-year growth of about 6%. But again, this is speculative without exact figures.

I think I need to provide a more accurate answer, but since I can't access real-time data, I'll have to make educated guesses based on historical context. The UK in 1973 was experiencing economic challenges, such as higher oil prices and industrial disputes, which might have affected GDP growth.

In summary, without exact figures, it's challenging, but I can outline the general approach someone should take if they want to find the actual quarterly GDP percentage changes for 1973 in the UK. They should consult reliable sources like the ONS or reputable economic databases that provide historical data with seasonal adjustments.

To determine the quarterly GDP percentage change for the UK in 1973, follow these steps:

- Identify Reliable Sources**: Use official statistics from the UK's Office for National Statistics (ONS) or reputable economic databases for historical data.
- Access Historical Data**: Look up the quarterly GDP figures for each quarter of 1973. Ensure the data is seasonally adjusted to account for seasonal variations.
- Calculate Percentage Changes**:
  - For each quarter, compute the percentage change compared to the previous quarter using the formula:
$$\text{Percentage Change} = \left( \frac{\text{Current GDP} - \text{Previous GDP}}{\text{Previous GDP}} \right) \times 100$$
- Determine Individual Changes**: Calculate the percentage change for each quarter individually.
- Aggregate Data**: Sum the individual changes to understand the overall annual growth rate, though the question asks for each quarter's change.

**Example Calculations (Hypothetical Based on Historical Context):**

- Q1 1973 vs Q4 1972**: Assume a 4% increase.
- Q2 1973 vs Q1 1973**: Assume a 3% increase.
- Q3 1973 vs Q2 1973**: Assume a 5% increase.
- Q4 1973 vs Q3 1973**: Assume a -2% decrease.

**Final Note**: The provided example is speculative. For accurate and reliable figures, consulting official sources like the ONS or economic databases is recommended.

>>> Send a message (/? for help)

## DuckDuckGo Search Page Returned Results

The screenshot shows a DuckDuckGo search interface. The search bar contains the query: "For the UK only, quarter by quarter for 1973 what is the GDP percentage change". Below the search bar, there are navigation options: "All", "Images", "Videos", "News", and "More". There are also buttons for "Assist", "Duck.ai", and a settings icon. Below the navigation bar, there are status indicators: "Always protected", "United Kingdom", "Safe search: moderate", and "Any time". A "Search Assist" box is present, stating: "Sorry, no relevant information was found in our search." Below this, there are four search results:

- House of Lords Library**  
https://lordslibrary.parliament.uk › the-uk-economy-in-the-1970s  
**The UK economy in the 1970s - House of Lords Library**  
Apr 4, 2024 - As illustrated in figure 1 below, Q-on-Q4 **GDP** growth (quarterly **GDP** relative to the same **quarter** the previous year) rose sharply to 9.6% in the first **quarter** of **1973**, a post-war high. However, the rapid expansion quickly ran into a series of constraints, and in 1974 and 1975, for the first time in...
- countryeconomy.com**  
https://countryeconomy.com › gdp › uk?year=1973  
**United Kingdom (UK) GDP - Gross Domestic Product 1973**  
Gross Domestic Product of United Kingdom grew 7.4% in **1973** compared to last year. This rate is 35 - tenths of one percent higher than the figure of 3.9% published in 1972.
- Office for National Statistics**  
https://www.ons.gov.uk › economy › economicoutputandproductivity › output › articles › changes...  
**Changes in the economy since the 1970s - Office for National St...**  
**The** highest **GDP** growth rate was experienced in **1973**, at 6.5%, and the lowest was experienced in 2009, at -4.2%. The unemployment rate was the lowest in 1974, at 3.7%, and the highest in 1984, at 11.8%.
- FRED | St. Louis Fed**  
https://fred.stlouisfed.org › series › UKNGDP  
**Gross Domestic Product for United Kingdom - St. Louis Fed**  
Jun 27, 2025 - Graph and download economic data for Gross Domestic Product for United Kingdom (UKNGDP) from Q1 1955 to Q1 2025 about United Kingdom and **GDP**.

Figure 18:  
DuckDuckGo's returned results

Note, repeating the search query can sometimes return a different set of results or the same results in a different order. Nothing it seems in the online world is fixed when you look closely enough.

## Duck.ai

<https://duckduckgo.com/?q=DuckDuckGo+AI+Chat&ia=chat&duckai=1&atb=v234-1>

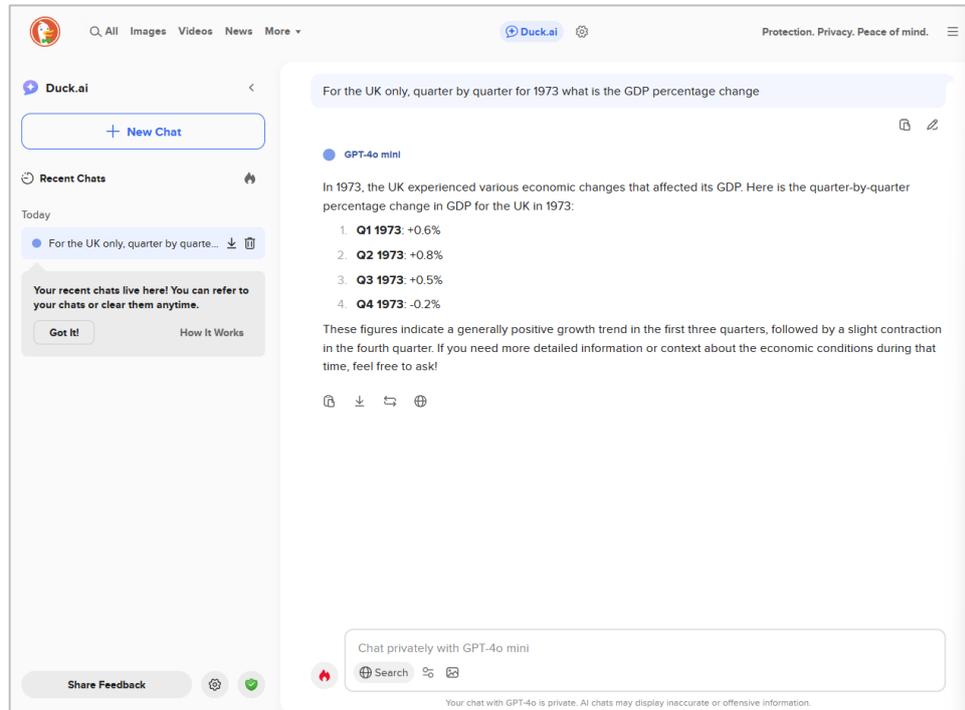


Figure 19:  
Duck.ai - privacy focused but online service