

02-Groq.com API for test set

In this notebook, we will make API calls to Groq.com to summarize the article (up to 20k characters) in the test set.

We would use Mixtral 8x7B as it has very good performance scores, and it's open source.

```
In [1]: # !pip install groq
```

```
In [2]: # Load api key
import json
import urllib.parse

with open('./data/credentials.json') as f:
    login = json.load(f)

api_key = login["GROQ_API_KEY"]
print(len(api_key))
```

56

```
In [3]: from groq import Groq

client = Groq(
    api_key=api_key,
)
```

```
In [4]: # test pause
import time
print("Hello")
time.sleep(2)
print("World")
```

Hello
World

```
In [5]: SLEEP_TIME = 10 # pause between requests
```

```
In [6]: def send_sumarize_request(content, model=client, min_words=250, max_words=500, quiet=False):
    """
    summarize the content
    input: context (text), model (groq_api client), max_words (int)
    output: summarized text
    """
    if not quiet:
        print("Sending request for text =", content[:100])

    result = ""
    prompt = f'Simplify and summarize in minimum {min_words} to maximum {max_words} words'
    try:
        completion = client.chat.completions.create(
            model="mixtral-8x7b-32768",
            prompt=prompt,
            max_tokens=max_words,
            min_tokens=min_words,
            temperature=0.1,
            top_p=0.95,
            stream=False,
        )
        result = completion.choices[0].text
    except Exception as e:
        print(f"Error: {e}")
    return result
```

```

        messages=[
            {
                "role": "user",
                "content": prompt
            },
        ],
        temperature=0.8,
        max_tokens=2048,
        top_p=1,
        stream=True,
        stop=None,
    )

    for chunk in completion:
        result += chunk.choices[0].delta.content or ""
    except Exception as err:
        print("Skipping, error : ", err)
        result = ""

    # pause to avoid hitting bandwidth limit (~ 14K token / minute)
    print(f"Pausing for {SLEEP_TIME} secs...", end="")
    time.sleep(SLEEP_TIME)
    print("OK")

    return result

```

In [7]: `import pandas as pd`

In [8]: `# Load in some data`

```

dev_df_filename = "../data/biolaysumm2024_data/eLife_test.jsonl"
df = pd.read_json(dev_df_filename,
                  orient="records",
                  lines=True
                )

df.head()

```

Out[8]:

	article	headings	keywords	id
0	Acylation of diverse carbohydrates occurs acro...	[Abstract, Introduction, Results and discussio...	[biochemistry and chemical biology, computatio...	elife-81547-v1
1	Honey bee ecology demands they make both rapid...	[Abstract, Introduction, Results, Discussion, ...	[computational and systems biology]	elife-86176-v2
2	Biguanides , including the world's most prescr...	[Abstract, Introduction, Results, Discussion, ...	[genetics and genomics]	elife-82210-v1
3	Ecological relationships between bacteria medi...	[Abstract, Introduction, Results, Discussion, ...	[microbiology and infectious disease, ecology]	elife-83152-v2
4	Gamma oscillations are believed to underlie co...	[Abstract, Introduction, Results, Discussion, ...	[neuroscience]	elife-83044-v2

```
In [9]: # apply to all rows in eval miniset
text_cap = 20_000 # temporarily limit to 20k characters because of API restriction

print("Summarization process started...")
df["groq_mistral_summary"] = df["article"].apply(lambda text: send_sumarize_request
print("Completed")
```

Summarization process started...

Sending request for text = Acylation of diverse carbohydrates occurs across all domains of life and can be catalysed by protein

Pausing for 10 secs...OK

Sending request for text = Honey bee ecology demands they make both rapid and accurate assessments of which flowers are most likely

Pausing for 10 secs...OK

Sending request for text = Biguanides, including the world's most prescribed drug for type 2 diabetes, metformin, not only lower

Pausing for 10 secs...OK

Sending request for text = Ecological relationships between bacteria mediate the services that gut microbiomes provide to their

Pausing for 10 secs...OK

Sending request for text = Gamma oscillations are believed to underlie cognitive processes by shaping the formation of transient

Pausing for 10 secs...OK

Sending request for text = Asynchronous replication of chromosome domains during S phase is essential for eukaryotic genome function

Pausing for 10 secs...OK

Sending request for text = While biological age in adults is often understood as representing general health and resilience, there is

Pausing for 10 secs...OK

Sending request for text = Many genes associated with asthma explain only a fraction of its heritability. Most genome-wide association

Pausing for 10 secs...OK

Sending request for text = By capitalising on positive biodiversity-productivity relationships, intercropping provides opportunities

Pausing for 10 secs...OK

Sending request for text = Many real-world decisions in social contexts are made while observing a partner's actions. To study

Pausing for 10 secs...OK

Sending request for text = An in vitro model of human ovarian follicles would greatly benefit the study of female reproduction

Pausing for 10 secs...OK

Sending request for text = Our interest in the genetic basis of skin color variation between populations led us to seek a Native

Pausing for 10 secs...OK

Sending request for text = Antioxidant intervention is considered to inhibit reactive oxygen species (ROS) and alleviate hyper

Pausing for 10 secs...OK

Sending request for text = The escape trajectory (ET) of prey - measured as the angle relative to the predator's approach path

Pausing for 10 secs...OK

Sending request for text = The recent development of prime editing (PE) genome engineering technologies has the potential to

Pausing for 10 secs...OK

Sending request for text = Multiple myeloma (MM) accounts for ~10% of all hematologic malignancies. Little is known about how

Pausing for 10 secs...OK

Sending request for text = Cas7-11 is a Type III-E CRISPR Cas effector that confers programmable RNA cleavage and has potential

Pausing for 10 secs...OK

Sending request for text = Addressing global biodiversity loss requires an expanded focus on multiple dimensions of biodiversity

Pausing for 10 secs...OK

Sending request for text = Single-molecule imaging provides a powerful way to study

biochemical processes in live cells , yet i
Pausing for 10 secs...OK
Sending request for text = The effect of calcium supplementation on bone mineral acc
retion in people under 35 years old is inco
Pausing for 10 secs...OK
Sending request for text = In biological systems , liquid and solid-like biomolecula
r condensates may contain the same molecule
Pausing for 10 secs...OK
Sending request for text = Kinesin superfamily proteins are microtubule-based molecu
lar motors driven by the energy of ATP hydr
Pausing for 10 secs...OK
Sending request for text = Highly potent animal stem cells either self renew or laun
ch complex differentiation programs , using
Pausing for 10 secs...OK
Sending request for text = Biological rhythms are ubiquitous . They can be generated
by circadian oscillators , which produce d
Pausing for 10 secs...OK
Sending request for text = Hydrogen peroxide is the most common reactive chemical th
at organisms face on the microbial battlefi
Pausing for 10 secs...OK
Sending request for text = Class switch recombination generates distinct antibody is
otypes critical to a robust adaptive immune
Pausing for 10 secs...OK
Sending request for text = Large-scale populations in the world have been vaccinated
with COVID-19 vaccines , however , breakth
Pausing for 10 secs...OK
Sending request for text = Plasmodium falciparum , the causative agent of malaria ,
remains a global health threat as parasites
Pausing for 10 secs...OK
Sending request for text = Early hematopoiesis is a continuous process in which hema
topoietic stem and progenitor cells (HSPCs
Pausing for 10 secs...OK
Sending request for text = Animals must learn to ignore stimuli that are irrelevant
to survival and attend to ones that enhance
Pausing for 10 secs...OK
Sending request for text = The specific functional properties of a tissue are distri
buted amongst its component cell types . Th
Pausing for 10 secs...OK
Sending request for text = Cardiovascular disease is the leading cause of death worl
dwide due to the inability of adult heart t
Pausing for 10 secs...OK
Sending request for text = The risk of second primary cancers (SPC) is increasing
after the first primary cancers (FPC) are
Pausing for 10 secs...OK
Sending request for text = In economic decision-making individuals choose between it
ems based on their perceived value . For bo
Pausing for 10 secs...OK
Sending request for text = Most phytophagous insect species exhibit a limited diet b
readth and specialize on a few or a single
Pausing for 10 secs...OK
Sending request for text = Multiple genetic changes in the enteric pathogen Yersinia
pseudotuberculosis have driven the emergen
Pausing for 10 secs...OK
Sending request for text = The peer review process is a critical step in ensuring th
e quality of scientific research . However
Pausing for 10 secs...OK

Sending request for text = Hepatic metastases are a poor prognostic factor of colorectal carcinoma (CRC) and new strategies t
Pausing for 10 secs...OK
Sending request for text = Although vaccination is broadly used in North American swine breeding herds , managing swine influen
Pausing for 10 secs...OK
Sending request for text = Aversive stimuli can cause hippocampal place cells to remap their firing fields , but it is not know
Pausing for 10 secs...OK
Sending request for text = HIV-1 infection results in the activation of inflammasome that may facilitate viral spread and estab
Pausing for 10 secs...OK
Sending request for text = The loss of a spouse is often cited as the most traumatic event in a person's life . However , for m
Pausing for 10 secs...OK
Sending request for text = Motor variability is a fundamental feature of developing systems allowing motor exploration and lear
Pausing for 10 secs...OK
Sending request for text = The Global Typhoid Genomics Consortium was established to bring together the typhoid research commun
Pausing for 10 secs...OK
Sending request for text = To synchronize flowering time with spring , many plants undergo vernalization , a floral-promotion p
Pausing for 10 secs...OK
Sending request for text = HIV-1 reservoir cells that circulate in peripheral blood during suppressive antiretroviral therapy (
Pausing for 10 secs...OK
Sending request for text = Circadian clocks are highly conserved transcriptional regulators that control ~24 hr oscillations in
Pausing for 10 secs...OK
Sending request for text = Membrane trafficking pathways perform important roles in establishing and maintaining the endosomal
Pausing for 10 secs...OK
Sending request for text = The effects of temperature on interaction strengths are important for understanding and forecasting
Pausing for 10 secs...OK
Sending request for text = The avian palaeognath phylogeny has been recently revised significantly due to the advancement of ge
Pausing for 10 secs...OK
Sending request for text = Ultrasonic vocalizations (USVs) fulfill an important role in communication and navigation in many
Pausing for 10 secs...OK
Sending request for text = Sarcomeres are the force-producing units of all striated muscles . Their nanoarchitecture critically
Pausing for 10 secs...OK
Sending request for text = Numerous cell functions are accompanied by phenotypic changes in viscoelastic properties , and measu
Pausing for 10 secs...OK
Sending request for text = Cardiac muscle has the highest mitochondrial density of any human tissue , but mitochondrial dysfunc
Pausing for 10 secs...OK
Sending request for text = Several previous lines of research have suggested , indirectly , that mouse lifespan is particularly
Pausing for 10 secs...OK
Sending request for text = Designer receptors exclusively activated by designer drugs (DREADDs) are chemogenetic tools for re

Pausing for 10 secs...OK
Sending request for text = Microorganism sensing of and responding to ambient chemical gradients regulates a myriad of microbial
Pausing for 10 secs...OK
Sending request for text = While neural crest development is known to be transcriptionally controlled via sequential activation
Pausing for 10 secs...OK
Sending request for text = Breathing , and the motor circuits that control it , is essential for life . At the core of respiration
Pausing for 10 secs...OK
Sending request for text = The powerful allure of social media platforms has been attributed to the human need for social reward
Pausing for 10 secs...OK
Sending request for text = The disassembly of microtubules can generate force and drive intracellular motility . During mitosis
Pausing for 10 secs...OK
Sending request for text = When attacked , hagfishes produce a soft , fibrous defensive slime within a fraction of a second by
Pausing for 10 secs...OK
Sending request for text = There are no pharmacological disease-modifying treatments with an enduring effect to mitigate the severity
Pausing for 10 secs...OK
Sending request for text = Fragility fracture is associated with an increased risk of mortality , but mortality is not part of
Pausing for 10 secs...OK
Sending request for text = Accurate brain tissue extraction on magnetic resonance imaging (MRI) data is crucial for analyzing
Pausing for 10 secs...OK
Sending request for text = Heme can serve as iron source in many environments , including the iron-poor animal host environment
Pausing for 10 secs...OK
Sending request for text = The seventh pandemic of the diarrheal cholera disease , which began in 1960 , is caused by the Gram-negative
Pausing for 10 secs...OK
Sending request for text = The hippocampus supports multiple cognitive functions including episodic memory . Recent work has highlighted
Pausing for 10 secs...OK
Sending request for text = Stimulator of interferon genes (STING) is activated after detection of cytoplasmic dsDNA by cGAS (Cardifin)
Pausing for 10 secs...OK
Sending request for text = Despite the remarkable successes of cancer immunotherapies , the majority of patients will experience
Pausing for 10 secs...OK
Sending request for text = Theory of Mind (ToM) refers to the cognitive ability to attribute mental states to other individuals
Pausing for 10 secs...OK
Sending request for text = Coordinated regulation of gene activity by transcriptional and translational mechanisms poise stem cells
Pausing for 10 secs...OK
Sending request for text = Billions of apoptotic cells are removed daily in a human adult by professional phagocytes (e . g . macrophages)
Pausing for 10 secs...OK
Sending request for text = Elaborate behaviours are produced by tightly controlled flexor-extensor motor neuron activation patterns
Pausing for 10 secs...OK
Sending request for text = Respiration can induce motion and CO2 fluctuation during

resting-state fMRI (rsfMRI) scans , which
Pausing for 10 secs...OK
Sending request for text = Predators that target multiple prey types are predicted to switch foraging modes according to prey p
Pausing for 10 secs...OK
Sending request for text = The evolution of human right-handedness has been intensively debated for decades . Manual lateraliza
Pausing for 10 secs...OK
Sending request for text = Williams-Beuren syndrome (WBS) is a rare genetic multisystemic disorder characterized by mild-to-m
Pausing for 10 secs...OK
Sending request for text = During development , the growing organism transits through a series of temporally regulated morpholo
Pausing for 10 secs...OK
Sending request for text = Cycling of co-substrates , whereby a metabolite is converted among alternate forms via different rea
Pausing for 10 secs...OK
Sending request for text = The origin and differentiation mechanism of articular chondrocytes remain poorly understood . Broadl
Pausing for 10 secs...OK
Sending request for text = Photosystem II (PSII) uses the energy from red light to split water and reduce quinone , an energy
Pausing for 10 secs...OK
Sending request for text = The skeletal system contains a series of sophisticated cellular lineages arising from the mesenchyma
Pausing for 10 secs...OK
Sending request for text = Organisms can adapt to an environment by taking multiple mutational paths . This redundancy at the g
Pausing for 10 secs...OK
Sending request for text = CD133 (prominin 1) is widely viewed as a cancer stem cell marker in association with drug resistan
Pausing for 10 secs...OK
Sending request for text = Ciliopathies manifest from sensory abnormalities to syndromic disorders with multi-organ pathologies
Pausing for 10 secs...OK
Sending request for text = While foraging for nectar and pollen , bees are exposed to a myriad of xenobiotics , including plant
Pausing for 10 secs...OK
Sending request for text = Candida albicans causes severe invasive candidiasis . C . albicans infection requires the virulence
Pausing for 10 secs...OK
Sending request for text = Tailoring interventions to patient subgroups can improve intervention outcomes for various condition
Pausing for 10 secs...OK
Sending request for text = Shared lineage has diverse effects on patterns of neuronal connectivity . In mammalian cortex , exci
Pausing for 10 secs...OK
Sending request for text = Healthy neuronal networks rely on homeostatic plasticity to maintain stable firing rates despite cha
Pausing for 10 secs...OK
Sending request for text = The tips of the developing respiratory buds are home to important progenitor cells marked by the exp
Pausing for 10 secs...OK
Sending request for text = Working memory (WM) is one of the most affected cognitive domains in multiple sclerosis (MS) , w
Pausing for 10 secs...OK

Sending request for text = Nociception is a neural process that animals have developed to avoid potentially tissue-damaging stimuli
Pausing for 10 secs...OK

Sending request for text = Plants often generate secondary metabolites as defense mechanisms against parasites . Although some
Pausing for 10 secs...OK

Sending request for text = Psychophysical methods are a cornerstone of psychology , cognitive science , and neuroscience where
Pausing for 10 secs...OK

Sending request for text = Models based on normative principles have played a major role in our understanding of how the brain
Pausing for 10 secs...OK

Sending request for text = Despite efforts from scientists and regulators , biodiversity is declining at an alarming rate . Unl
Pausing for 10 secs...OK

Sending request for text = When interacting with the dynamic world , the brain receives outdated sensory information , due to t
Pausing for 10 secs...OK

Sending request for text = One signature of the human brain is its ability to derive knowledge from language inputs , in addition
Pausing for 10 secs...OK

Sending request for text = Replay , the sequential reactivation within a neuronal ensemble , is a central hippocampal mechanism
Pausing for 10 secs...OK

Sending request for text = Impaired spermatogenesis and male infertility are common manifestations associated with mitochondria
Pausing for 10 secs...OK

Sending request for text = Analogues of the hepatokine fibroblast growth factor 21 (FGF21) are in clinical development for type 2 diabetes
Pausing for 10 secs...OK

Sending request for text = The neurophysiology of cells and tissues are monitored electrophysiologically and optically in diverse
Pausing for 10 secs...OK

Sending request for text = Infection with Influenza A virus (IAV) causes the well-known symptoms of the flu , including fever
Pausing for 10 secs...OK

Sending request for text = Apical-basal cell polarity loss is a founding event in epithelial-mesenchymal transition and epithelial
Pausing for 10 secs...OK

Sending request for text = Humans learn about the environment either directly by interacting with it or indirectly by seeking information
Pausing for 10 secs...OK

Sending request for text = Synaptic communication relies on the fusion of synaptic vesicles with the plasma membrane , which leads to
Pausing for 10 secs...OK

Sending request for text = Sauropsidia was a taxonomically and ecomorphologically diverse clade of Mesozoic marine reptiles
Pausing for 10 secs...OK

Sending request for text = Smoking-associated DNA methylation levels identified through epigenome-wide association studies (EWAS)
Pausing for 10 secs...OK

Sending request for text = The structure and function of the vertebrate retina have been extensively studied across species with
Pausing for 10 secs...OK

Sending request for text = Fixing cells with paraformaldehyde (PFA) is an essential step in numerous biological techniques as

Pausing for 10 secs...OK
Sending request for text = Uropathogenic Escherichia coli (UPEC) proliferate within superficial bladder umbrella cells to form
Pausing for 10 secs...OK
Sending request for text = Ammonium (NH_4^+) is essential to generate the nitrogenous building blocks of life . It gets assimilated
Pausing for 10 secs...OK
Sending request for text = Reef-building corals thrive in oligotrophic environments due to their possession of endosymbiotic algae
Pausing for 10 secs...OK
Sending request for text = Nephronophthisis (NPHP) is a ciliopathy characterized by renal fibrosis and cyst formation , and associated with
Pausing for 10 secs...OK
Sending request for text = The extinct Steller's sea cow (Hydrodamalis gigas; †1768) was a whale-sized marine mammal that man
Pausing for 10 secs...OK
Sending request for text = Expression of activated Ras , RasV12 , provides Drosophila cultured cells with a proliferation and survival
Pausing for 10 secs...OK
Sending request for text = Multiciliated cells (MCCs) are terminally differentiated epithelia that assemble multiple motile cilia
Pausing for 10 secs...OK
Sending request for text = The human endometrium experiences repetitive cycles of tissue wounding characterised by piecemeal shedding
Pausing for 10 secs...OK
Sending request for text = Adaptation to fluctuating environmental conditions is difficult to achieve . Phase variation mechanisms
Pausing for 10 secs...OK
Sending request for text = The rate and accuracy of translation hinges upon multiple components - including transfer RNA (tRNA)
Pausing for 10 secs...OK
Sending request for text = The COVID-19 pandemic has resulted in a step change in the scale of sequencing data , with more genomes
Pausing for 10 secs...OK
Sending request for text = RNA molecules are localized to specific subcellular regions through interactions between RNA regulatory
Pausing for 10 secs...OK
Sending request for text = Aging is often associated with a loss of cell type identity that results in an increase in transcription
Pausing for 10 secs...OK
Sending request for text = Tumour heterogeneity is thought to be a major barrier to successful cancer treatment due to the presence
Pausing for 10 secs...OK
Sending request for text = DMRT1 is the testis-determining factor in several species of vertebrates , but its involvement in male
Pausing for 10 secs...OK
Sending request for text = Obesity induced by high-fat diet (HFD) is a multifactorial disease including genetic , physiological
Pausing for 10 secs...OK
Sending request for text = During vertebrate embryogenesis , the germ layers are patterned by secreted Nodal signals . In the case
Pausing for 10 secs...OK
Sending request for text = Type I interferon (IFN-I) -mediated antiviral responses are central to host defense against viral
Pausing for 10 secs...OK
Sending request for text = Microtubules (MTs) are built from α -/ β -tubulin dimers and

nd used as tracks by kinesin and dynein mot
Pausing for 10 secs...OK
Sending request for text = The development of multiple chromosome-scale reference ge
nome sequences in many taxonomic groups has
Pausing for 10 secs...OK
Sending request for text = Most bacteria exist and interact within polymicrobial com
munities . These interactions produce uniqu
Pausing for 10 secs...OK
Sending request for text = The quantity and complexity of data being generated and p
ublished in biology has increased substanti
Pausing for 10 secs...OK
Sending request for text = Sequence variation in enhancers that control cell-type-sp
ecific gene transcription contributes signi
Pausing for 10 secs...OK
Sending request for text = Viruses interact with the intracellular transport machine
ry to promote viral replication . Such host
Pausing for 10 secs...OK
Sending request for text = Actin isoforms organize into distinct networks that are e
ssential for the normal function of eukaryo
Pausing for 10 secs...OK
Sending request for text = High-throughput transgenesis using synthetic DNA librarie
s is a powerful method for systematically e
Pausing for 10 secs...OK
Sending request for text = To reach their final destinations , outer membrane protei
ns (OMPs) of gram-negative bacteria under
Pausing for 10 secs...OK
Sending request for text = Midbrain dopamine (DA) neurons are key regulators of ba
sal ganglia functions . The axonal domain o
Pausing for 10 secs...OK
Sending request for text = Mobile health (mHealth) interventions , which require o
wnership of mobile phones , are being inves
Pausing for 10 secs...OK
Sending request for text = Severe acute respiratory syndrome coronavirus 2 (SARS-Co
V-2) antibody levels can be used to assess
Pausing for 10 secs...OK
Completed

In [10]: `df.head()`

Out[10]:

	article	headings	keywords	id	groq_mistral_summary
0	Acylation of diverse carbohydrates occurs acro...	[Abstract, Introduction, Results and discussio...	[biochemistry and chemical biology, computatio...	elife-81547-v1	Acyltransferase family 3 (AT3) domain-containi...
1	Honey bee ecology demands they make both rapid...	[Abstract, Introduction, Results, Discussion, ...	[computational and systems biology]	elife-86176-v2	Honey bees make rapid and accurate decisions w...
2	Biguanides , including the world's most prescr...	[Abstract, Introduction, Results, Discussion, ...	[genetics and genomics]	elife-82210-v1	Metformin, a biguanide and first-line treatmen...
3	Ecological relationships between bacteria medi...	[Abstract, Introduction, Results, Discussion, ...	[microbiology and infectious disease, ecology]	elife-83152-v2	In this study, researchers aimed to understand...
4	Gamma oscillations are believed to underlie co...	[Abstract, Introduction, Results, Discussion, ...	[neuroscience]	elife-83044-v2	Gamma oscillations in the brain are believed t...

```
In [11]: # check how many rows have blank result (some errors)
empty_df = df.query("groq_mistral_summary.str.strip() == ''")
print(len(empty_df))
```

0

```
In [12]: # attempt to retry for blank rows (due to some errors)
retry = True
if retry:
    print("Retrying for empty results...")
    for i in range(len(df)):
        item = df.iloc[i]
        if item["groq_mistral_summary"] == "":
            print("Item =", i)
            text = item["article"]
            df.at[i, "groq_mistral_summary"] = send_sumarize_request(text[:text_cap
            print("Completed")
```

Retrying for empty results...

Completed

```
In [13]: output_path = "./data/output/test_set/"
output_filename = "elife_groq_mistral_summary.csv"

print("Writing to file ", output_filename)
df.to_csv(output_path+output_filename,
          index = False)
```

```
)  
print("Completed")
```

Writing to file elife_groq_mistral_summary.csv
Completed

```
In [14]: output_filename = "elife_groq_mistral_summary.json"  
  
print("Writing to file ", output_filename)  
df.to_json(output_path+output_filename,  
           orient="records",  
           )  
print("Completed")
```

Writing to file elife_groq_mistral_summary.json
Completed

```
In [15]: # process the PLOS dataset:  
dev_df_filename = "../data/biolaysumm2024_data/PLOS_test.jsonl"  
df = pd.read_json(dev_df_filename,  
                  orient="records",  
                  lines=True  
                  )  
df.head()
```

Out[15]:	article	headings	keywords	id
0	Lung-resident (LR) mesenchymal stem and stro...	[Abstract, Introduction, Results, Discussion, ...]	[immune system, medical conditions, molecular ...]	journal.ppat.1009789
1	Visceral leishmaniasis (VL) is endemic in So...	[Abstract, Introduction, Methods, Results, Dis...]	[neonates, clinical laboratory sciences, trans...]	journal.pntd.0007992
2	A high burden of Salmonella enterica subspecie...	[Abstract, Introduction, Methods, Results, Dis...]	[pathogens, medical conditions, taxonomy, bact...]	journal.pntd.0010704
3	Severe Acute Respiratory Syndrome Coronavirus-...	[Abstract, Introduction, Results, Discussion, ...]	[pathogens, amniotes, medical conditions, bind...]	journal.ppat.1010691
4	Many fungal species utilize hydroxyderivatives...	[Abstract, Introduction, Results and discussio...]	[taxonomy, proteins, chemistry, genetics, enzy...]	journal.pgen.1009815

```
In [16]: # apply to all rows in eval miniset  
text_cap = 20_000 # temporarily limit to 20k characters due to API restriction, se  
  
print("Summarization process started...")
```

```
df["groq_mistral_summary"] = df["article"].apply(lambda text: send_sumarize_request(text))  
print("Completed")
```

Summarization process started...

Sending request for text = Lung-resident (LR) mesenchymal stem and stromal cells (MSCs) are key elements of the alveolar ni

Pausing for 10 secs...OK

Sending request for text = Visceral leishmaniasis (VL) is endemic in South Sudan , where outbreaks occur frequently . Because

Pausing for 10 secs...OK

Sending request for text = A high burden of Salmonella enterica subspecies enterica serovar Typhi (S . Typhi) bacteremia has

Pausing for 10 secs...OK

Sending request for text = Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-Co V-2) marks the third novel β -coronavirus t

Pausing for 10 secs...OK

Sending request for text = Many fungal species utilize hydroxyderivatives of benzene and benzoic acid as carbon sources . The y

Pausing for 10 secs...OK

Sending request for text = Improving biological plausibility and functional capacity are two important goals for brain models t

Pausing for 10 secs...OK

Sending request for text = Trisomy of human chromosome 21 (HSA21) causes Down synd rome (DS) . The trisomy does not simply r

Pausing for 10 secs...OK

Sending request for text = Ongoing oscillations and evoked responses are two main ty pes of neuronal activity obtained with dive

Pausing for 10 secs...OK

Sending request for text = Crimean-Congo hemorrhagic fever virus (CCHFV) is a tick -borne virus that can cause a hemorrhagic f

Pausing for 10 secs...OK

Sending request for text = Working memory (WM) deficits have been widely documente d in schizophrenia (SZ) , and almost all

Pausing for 10 secs...OK

Sending request for text = The HIV-1 envelope glycoprotein (Env) mediates viral en try via conformational changes associated w

Pausing for 10 secs...OK

Sending request for text = Extensive amounts of multi-omics data and multiple cancer subtyping methods have been developed rapi

Pausing for 10 secs...OK

Sending request for text = Ticks (order: Ixodida) are a highly diverse and ecologi cally important group of ectoparasitic bloo

Pausing for 10 secs...OK

Sending request for text = Meiotic recombination is an essential biological process that ensures faithful chromosome segregatio

Pausing for 10 secs...OK

Sending request for text = The APOBEC3 (A3) genes encode cytidine deaminase protei ns with potent antiviral and anti-retroelem

Pausing for 10 secs...OK

Sending request for text = Variant allele frequencies (VAF) encode ongoing evoluti on and subclonal selection in growing tumou

Pausing for 10 secs...OK

Sending request for text = African trypanosomes are parasites mainly transmitted by tsetse flies . They cause trypanosomiasis i

Pausing for 10 secs...OK

Sending request for text = Hybridization is a fascinating evolutionary phenomenon th at raises the question of how species maint

Pausing for 10 secs...OK

Sending request for text = Human immunodeficiency virus (HIV) vaccines have not be

en successful in clinical trials . Dimeric
Pausing for 10 secs...OK
Sending request for text = Understanding the relationships between biological processes is paramount to unravel pathophysiology
Pausing for 10 secs...OK
Sending request for text = If a mother contracts the Zika Virus before or during pregnancy , then there is a risk of the child
Pausing for 10 secs...OK
Sending request for text = Two-component , self-assembling nanoparticles represent a versatile platform for multivalent presentation
Pausing for 10 secs...OK
Sending request for text = Pseudorabies virus (PRV) has evolved various immune evasion mechanisms that target host antiviral
Pausing for 10 secs...OK
Sending request for text = Arthropod-borne viruses (arboviruses) require replication across a wide range of temperatures to propagate
Pausing for 10 secs...OK
Sending request for text = Olfactory integration is important for survival in a natural habitat . However , how the nervous system
Pausing for 10 secs...OK
Sending request for text = Extracellular vesicles (EVs) are protein-loaded nanoscaled particles that are extracellularly released
Pausing for 10 secs...OK
Sending request for text = Mycetoma is considered a neglected tropical disease globally . However , data on its burden and the
Pausing for 10 secs...OK
Sending request for text = The aim of Malaysia to eliminate malaria nationwide by 2020 seems need to be prolonged . Whilst Malaysia
Pausing for 10 secs...OK
Sending request for text = Latent Epstein-Barr virus (EBV) infection is strongly associated with several malignancies , including
Pausing for 10 secs...OK
Sending request for text = Improving our understanding of the genes regulating grain yield can contribute to the development of
Pausing for 10 secs...OK
Sending request for text = Scrub typhus is an emerging vector-borne zoonosis , caused by Orientia spp . and transmitted by larval
Pausing for 10 secs...OK
Sending request for text = During the 2015–2017 Zika epidemic , dengue and chikungunya—two other viral diseases with the same vector
Pausing for 10 secs...OK
Sending request for text = HBV is an enveloped DNA virus that replicates its DNA genome via reverse transcription of a pregenomic
Pausing for 10 secs...OK
Sending request for text = Tuberculosis caused by Mycobacterium tuberculosis (Mtb) is a significant public health concern , especially
Pausing for 10 secs...OK
Sending request for text = The biological master clock , suprachiasmatic nucleus (SCN of rat and mouse) , is composed of ~10,000
Pausing for 10 secs...OK
Sending request for text = To characterize the functional role of the left-ventral occipito-temporal cortex (lvOT) during reading
Pausing for 10 secs...OK
Sending request for text = The circadian clock exerts significant influence on the immune system and disruption of circadian rhythms
Pausing for 10 secs...OK

Sending request for text = Rapid urbanization , global trade , and the exceptionally great numbers of worldwide visitors during
Pausing for 10 secs...OK

Sending request for text = Fat stored in the form of lipid droplets has long been considered a defining characteristic of cytop
Pausing for 10 secs...OK

Sending request for text = Neuronal damage is a major consequence of bacterial meningitis , but little is known about mechanism
Pausing for 10 secs...OK

Sending request for text = Conspecific male animals fight for resources such as food and mating opportunities but typically sto
Pausing for 10 secs...OK

Sending request for text = Mosquito viruses cause unpredictable outbreaks of disease . Recently , several unassigned viruses is
Pausing for 10 secs...OK

Sending request for text = mTOR , a serine/threonine protein kinase that is involved in a series of critical cellular processes
Pausing for 10 secs...OK

Sending request for text = As a common vector-borne disease , dengue fever remains challenging to predict due to large variatio
Pausing for 10 secs...OK

Sending request for text = Leprosy and leprosy-related stigma can have a major impact on psychosocial wellbeing of persons affe
Pausing for 10 secs...OK

Sending request for text = Correct bioriented attachment of sister chromatids to the mitotic spindle is essential for chromosom
Pausing for 10 secs...OK

Sending request for text = T cell epitope candidates are commonly identified using computational prediction tools in order to e
Pausing for 10 secs...OK

Sending request for text = Epidemiological research suggests that paternal obesity may increase the risk of fathering small for
Pausing for 10 secs...OK

Sending request for text = Several barriers protect the central nervous system (CNS) from pathogen invasion . Yet viral infec
Pausing for 10 secs...OK

Sending request for text = Potent and durable vaccine responses will be required for control of malaria caused by Plasmodium fa
Pausing for 10 secs...OK

Sending request for text = Lysine malonylation is a post-translational modification (PTM) , which regulates many cellular pro
Pausing for 10 secs...OK

Sending request for text = Nematode parasites infect approximately 1 . 5 billion people globally and are a significant public h
Pausing for 10 secs...OK

Sending request for text = HIV-1 infects CD4 T lymphocytes (CD4TL) through binding the chemokine receptors CCR5 or CXCR4 . CX
Pausing for 10 secs...OK

Sending request for text = Zoonotic cutaneous leishmaniasis (ZCL) is endemic in Palestine and transmitted by Phlebotomus sand
Pausing for 10 secs...OK

Sending request for text = Non-coding regions of viral RNA (vRNA) genomes are critically important in the regulation of gene
Pausing for 10 secs...OK

Sending request for text = Praziquantel (PZQ) is currently the only recommended drug for infection and disease caused by the

Pausing for 10 secs...OK
Sending request for text = Drug resistance in viruses represents one of the major challenges of healthcare . As part of an effort
Pausing for 10 secs...OK
Sending request for text = A group of four human inhabited Nancowry Islands in Nicobar district in the Andaman and Nicobar Islands
Pausing for 10 secs...OK
Sending request for text = CREB-binding protein (CBP) participates in numerous transcription events . However , cell-intrinsic
Pausing for 10 secs...OK
Sending request for text = We report the identification of two orthobunyaviruses , Melao virus (MELV) and Oropouche virus (O
Pausing for 10 secs...OK
Sending request for text = The development of efficient vaccines against COVID-19 is an emergent need for global public health
Pausing for 10 secs...OK
Sending request for text = Morbidity management and disability prevention (MMDP) services are essential for the management of
Pausing for 10 secs...OK
Sending request for text = The Human T-cell leukemia virus type 1 (HTLV-1) orf I-encoded accessory protein p8 is cleaved from
Pausing for 10 secs...OK
Sending request for text = Sensory deprivation has long been known to cause hallucinations or “phantom” sensations , the most common
Pausing for 10 secs...OK
Sending request for text = Porcine hemagglutinating encephalomyelitis virus (PHEV) is a highly neurotropic coronavirus belonging
Pausing for 10 secs...OK
Sending request for text = The regulation of glycometabolism homeostasis is vital to maintain health and development of animal
Pausing for 10 secs...OK
Sending request for text = The primary immunological target of COVID-19 vaccines is the SARS-CoV-2 spike (S) protein . S is encoded
Pausing for 10 secs...OK
Sending request for text = Impaired type I interferons (IFNs) production or signaling have been associated with severe COVID-19
Pausing for 10 secs...OK
Sending request for text = Freshwater snails of the genera Biomphalaria , Bulinus , and Oncomelania are intermediate hosts of schistosomes
Pausing for 10 secs...OK
Sending request for text = Food and water-borne illness caused by ingestion of (oocysts) of Cryptosporidium and Giardia is one of the leading causes of
Pausing for 10 secs...OK
Sending request for text = Echinococcosis is a zoonotic parasitic disease caused by larval stages of cestodes belonging to the family Taeniidae
Pausing for 10 secs...OK
Sending request for text = Network-based intervention strategies can be effective and cost-efficient approaches to curtailing human
Pausing for 10 secs...OK
Sending request for text = Extracellular recording is an accessible technique used in animals and humans to study the brain physiology
Pausing for 10 secs...OK
Sending request for text = Co-infection of the four major species of human malaria parasite Plasmodium falciparum (Pf) , P. vivax , P. malariae , and P. knowlesi
Pausing for 10 secs...OK
Sending request for text = Effective and powerful survival mediation models are currently

ently lacking . To partly fill such knowled
Pausing for 10 secs...OK
Sending request for text = Environment shapes development through a phenomenon calle
d developmental plasticity . Deciphering it
Pausing for 10 secs...OK
Sending request for text = Mitochondria are vital organelles inside the cell and con
tribute to intracellular calcium (Ca^{2+}) d
Pausing for 10 secs...OK
Sending request for text = Hemorrhagic fever with renal syndrome (HFRS) is a zoono
sis characterized by clinical features of h
Pausing for 10 secs...OK
Sending request for text = Helicobacter pylori chronically infects the stomach of ap
proximately half of the world's population
Pausing for 10 secs...OK
Sending request for text = Soil-transmitted helminth (STH) infections cause signif
icant morbidity in children and women of re
Pausing for 10 secs...OK
Sending request for text = Chikungunya virus (CHIKV) is a re-emerging mosquito-bor
ne alphavirus , which has rapidly spread ar
Pausing for 10 secs...OK
Sending request for text = In order to adjust to changing environmental conditions ,
bacteria use nucleotide second messengers
Pausing for 10 secs...OK
Sending request for text = Epstein-Barr virus (EBV) causes lymphomas and epithelia
l cell cancers . Though generally silent in
Pausing for 10 secs...OK
Sending request for text = Buruli ulcer is one of the 20 neglected tropical diseases
in the world . This necrotizing hypodermi
Pausing for 10 secs...OK
Sending request for text = The co-localization of Cluster-of-Differentiation-44 prot
ein (CD44) and cytoplasmic adaptors in sp
Pausing for 10 secs...OK
Sending request for text = Collective living systems regularly achieve cooperative e
mergent functions that individual organisms
Pausing for 10 secs...OK
Sending request for text = Culex pipiens quinquefasciatus Say (Cx . quinquefasciatu
s) and Culex pipiens form molestus Forskal
Pausing for 10 secs...OK
Sending request for text = During the exploration of novel environments , place fiel
ds are rapidly formed in hippocampal CA1 ne
Pausing for 10 secs...OK
Sending request for text = Molecular xenomonitoring (MX) , the detection of parasi
te nucleic acid in the vector population ,
Pausing for 10 secs...OK
Sending request for text = Deposition of complement factors on Mycobacterium leprae
may enhance phagocytosis . Such deposition
Pausing for 10 secs...OK
Sending request for text = Nongenetic phenotypic variation can either speed up or sl
ow down adaptive evolution . We show that i
Pausing for 10 secs...OK
Sending request for text = Arthropod-borne viruses infect both mosquito and mammalia
n hosts . While much is known about virus-h
Pausing for 10 secs...OK
Sending request for text = Temporal environmental variations affect diversity in com
munities of competing populations . In part
Pausing for 10 secs...OK

Sending request for text = Functional non-coding (fnc) RNAs are nucleotide sequences of varied lengths , structures , and mechanisms
Pausing for 10 secs...OK

Sending request for text = We investigate the source detection problem in epidemiology , which is one of the most important issues
Pausing for 10 secs...OK

Sending request for text = Recent state-of-the-art multiplex imaging techniques have expanded the depth of information that can be obtained
Pausing for 10 secs...OK

Sending request for text = Leptospirosis has been described as a biphasic disease consisting of hematogenous dissemination to meninges
Pausing for 10 secs...OK

Sending request for text = Talaromycosis is a serious regional disease endemic in Southeast Asia . In China , *Talaromyces marneffei* is the main pathogen
Pausing for 10 secs...OK

Sending request for text = Human toxocariasis has been identified as an under-diagnosed parasitic zoonosis and health disparity
Pausing for 10 secs...OK

Sending request for text = Envenoming by scorpions in genus *Tityus* is a public health problem in Tropical America . One of the main causes of mortality is respiratory failure
Pausing for 10 secs...OK

Sending request for text = Bat-pollinated flowers have to attract their pollinators in absence of light and therefore some species have evolved unique strategies
Pausing for 10 secs...OK

Sending request for text = Elimination of the intermediate snail host of *Schistosoma* is the most effective way to control schistosomiasis
Pausing for 10 secs...OK

Sending request for text = *Giardia duodenalis* causes giardiasis , a major diarrheal disease in humans worldwide whose treatment is challenging
Pausing for 10 secs...OK

Sending request for text = The assimilation , incorporation , and metabolism of sulfur is a fundamental process across all domains of life
Pausing for 10 secs...OK

Sending request for text = Juvenile hormone (JH) acts as a gonadotrophic hormone stimulating insect vitellogenesis and oogenesis
Pausing for 10 secs...OK

Sending request for text = Alcoholic-related liver disease (ALD) is the cause of more than half of all liver-related deaths . The pathogenesis is complex
Pausing for 10 secs...OK

Sending request for text = To remodel functional neuronal connectivity , neurons often alter dendrite arbors through elimination and growth
Pausing for 10 secs...OK

Sending request for text = Vγ9Vδ2 T cells rapidly respond to phosphoantigens produced by *Plasmodium falciparum* in an innate-like manner
Pausing for 10 secs...OK

Sending request for text = Owing to their plasticity , intrinsically disordered and multidomain proteins require descriptions beyond static structures
Pausing for 10 secs...OK

Sending request for text = Enterotoxigenic *Escherichia coli* (ETEC) is one of the top aetiological agents of diarrhea in children
Pausing for 10 secs...OK

Sending request for text = Longitudinal intervention studies with repeated measurements over time are an important type of experimental design
Pausing for 10 secs...OK

Sending request for text = We study the time evolution of symptoms (signs) with some defects in the dynamics of a reaction network
Pausing for 10 secs...OK

Pausing for 10 secs...OK
Sending request for text = The relationship between transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-
Pausing for 10 secs...OK
Sending request for text = Counting the number of species , items , or genes that are shared between two groups , sets , or communities
Pausing for 10 secs...OK
Sending request for text = To determine the diagnostic validity of the enzyme-linked immunosorbent assay (ELISA) and Rapid Diagnostic
Pausing for 10 secs...OK
Sending request for text = Computational modeling of dopamine transmission is challenged by complex underlying mechanisms . Here
Pausing for 10 secs...OK
Sending request for text = Female tick salivary glands undergo rapid degeneration several days post engorgement . This degeneration
Pausing for 10 secs...OK
Sending request for text = In the last 20 years yellow fever (YF) has seen dramatic changes to its incidence and geographic distribution
Pausing for 10 secs...OK
Sending request for text = Obsessive-compulsive disorder (OCD) is characterized by uncontrollable repetitive actions and thoughts
Pausing for 10 secs...OK
Sending request for text = The Apicomplexa phylum comprises thousands of distinct intracellular parasite species , including ciliates
Pausing for 10 secs...OK
Sending request for text = One of the unique traits of membrane proteins is that a significant fraction of their hydrophobic amino acids
Pausing for 10 secs...OK
Sending request for text = Pandemic management requires reliable and efficient dynamical simulation to predict and control disease
Pausing for 10 secs...OK
Sending request for text = A major limitation to understanding the etiopathogenesis of environmental enteric dysfunction (EED)
Pausing for 10 secs...OK
Sending request for text = Real-world agents , humans as well as animals , observe each other during interactions and choose their
Pausing for 10 secs...OK
Sending request for text = Telomerase activity is the principal telomere maintenance mechanism in human cancers , however 15% of
Pausing for 10 secs...OK
Sending request for text = Toxoplasma gondii possesses an armada of secreted virulent factors that enable parasite invasion and
Pausing for 10 secs...OK
Sending request for text = The stimulation of a single neuron in the rat somatosensory cortex can elicit a behavioral response
Pausing for 10 secs...OK
Sending request for text = Human skin is continuously exposed to environmental DNA damage leading to the accumulation of somatic
Pausing for 10 secs...OK
Sending request for text = Clonorchiasis , caused by the fish-borne trematode Clonorchis sinensis , is a neglected tropical disease
Pausing for 10 secs...OK
Sending request for text = Skin problems cause significant sickness in communities with poor living conditions , but they have
Pausing for 10 secs...OK
Sending request for text = Pathogenic mycobacteria actively dysregulate protective host

ost immune signalling pathways during infec
Pausing for 10 secs...OK
Sending request for text = Robustness in developing and homeostatic tissues is supported by various types of spatiotemporal cel
Pausing for 10 secs...OK
Sending request for text = Existing studies have demonstrated that dysregulation of microRNAs (miRNAs or miRs) is involved in
Pausing for 10 secs...OK
Sending request for text = Synaptic efficacy is subjected to activity-dependent changes on short- and long time scales . While
Pausing for 10 secs...OK
Sending request for text = Bats are the natural reservoir host for a number of zoonotic viruses , including Hendra virus (HeV
Pausing for 10 secs...OK
Sending request for text = Evidence from both GWAS and clinical observation has suggested that certain psychiatric , metabolic
Pausing for 10 secs...OK
Sending request for text = Cell culture remains as the golden standard for primary isolation of viruses in clinical specimens .
Pausing for 10 secs...OK
Sending request for text = Herpesviruses are known to encode a number of inhibitors of host cell death , including RIP Homotypi
Pausing for 10 secs...OK
Sending request for text = Infectious disease risk is driven by three interrelated components: exposure , hazard , and vulnerab
Pausing for 10 secs...OK
Sending request for text = Nigeria continues to experience ever increasing annual outbreaks of Lassa fever (LF) . The World H
Pausing for 10 secs...OK
Sending request for text = Schistosomes are blood dwelling parasites that affect more than 260 million people globally , and ov
Pausing for 10 secs...OK
Sending request for text = Antibodies targeting the hepatitis C virus (HCV) envelope glycoprotein E2 are associated with dela
Pausing for 10 secs...OK
Completed

In [17]: `df.head()`

Out[17]:

	article	headings	keywords	id	groq_mistral_summary
0	Lung-resident (LR) mesenchymal stem and stro...	[Abstract, Introduction, Results, Discussion, ...	[immune system, medical conditions, molecular ...	journal.ppat.1009789	Lung-resident mesenchymal stem and stromal cel...
1	Visceral leishmaniasis (VL) is endemic in So...	[Abstract, Introduction, Methods, Results, Dis...	[neonates, clinical laboratory sciences, trans...	journal.pntd.0007992	Visceral leishmaniasis (VL) is a protozoan dis...
2	A high burden of Salmonella enterica subspecie...	[Abstract, Introduction, Methods, Results, Dis...	[pathogens, medical conditions, taxonomy, bact...	journal.pntd.0010704	A study was conducted to genetically character...
3	Severe Acute Respiratory Syndrome Coronavirus-...	[Abstract, Introduction, Results, Discussion, ...	[pathogens, amniotes, medical conditions, bind...	journal.ppat.1010691	The Severe Acute Respiratory Syndrome Coronavi...
4	Many fungal species utilize hydroxyderivatives...	[Abstract, Introduction, Results and discussio...	[taxonomy, proteins, chemistry, genetics, enzy...	journal.pgen.1009815	The yeast Candida parapsilosis metabolizes hyd...

```
In [18]: # check how many rows have blank result (some errors)
empty_df = df.query("groq_mistral_summary.str.strip() == ''")
print(len(empty_df))
```

0

```
In [19]: # attempt to retry for blank rows (due to some errors)
retry = True
if retry:
    print("Retrying for empty results...")
    for i in range(len(df)):
        item = df.iloc[i]
        if item["groq_mistral_summary"] == "":
            print("Item =", i)
            text = item["article"]
            df.at[i, "groq_mistral_summary"] = send_sumarize_request(text[:text_cap
            print("Completed")
```

Retrying for empty results...

Completed

```
In [20]: output_path = "./data/output/test_set/"
output_filename = "plos_groq_mistral_summary.csv"
```

```
print("Writing to file ", output_filename)
df.to_csv(output_path+output_filename,
          index = False
          )
print("Completed")
```

Writing to file plos_groq_mistral_summary.csv
Completed

In [21]: output_filename = "plos_groq_mistral_summary.json"

```
print("Writing to file ", output_filename)
df.to_json(output_path+output_filename,
           orient="records",
           )
print("Completed")
```

Writing to file plos_groq_mistral_summary.json
Completed

In []: