Unstructured Loaders

LangChain uses Chroma as the vectorstore to index and search embeddings

In [ ]:
```
! pip install Chroma
```

In [ ]:
```
! pip install chromadb
```

This example showcases question answering over documents. We have chosen this as the example for getting started because it nicely combines a lot of different elements (Text splitters, embeddings, vectorstores) and then also shows how to use them in a chain.

Question answering over documents consists of four steps:

- Create an index
- Create a Retriever from that index
- Create a question answering chain
- Ask questions!

First, let's import some common classes we'll use no matter what.

In [ ]:
```
!pip install langchain
```

In [ ]:
```
from langchain.chains import RetrievalQA
from langchain.llms import OpenAI
```

Next in the generic setup, let's specify the document loader we want to use. You can download the state_of_the_union.txt file here

Text Documents

In [181…
```
from langchain.document_loaders import TextLoader
loader = TextLoader("../input/state_of_the_union.txt",encoding='utf-8')
```

Create index over document

In [ ]:
```
!pip install openai
```

In [ ]:
```
!pip install tiktoken
```

In [182…
```
from langchain.indexes import VectorstoreIndexCreator
index=VectorstoreIndexCreator().from_loaders([loader])
```

```
Using embedded DuckDB without persistence: data will be transient
```

It is possible to configure the index creation this way

In [183…
```
from langchain.vectorstores import Chroma
from langchain.embeddings import OpenAIEmbeddings
from langchain.text_splitter import CharacterTextSplitter


index_creator = VectorstoreIndexCreator(
    vectorstore_cls=Chroma,
    embedding=OpenAIEmbeddings(),
    text_splitter=CharacterTextSplitter(chunk_size=1000, chunk_overlap=0)
)
```

Querying the index

In [184…
```
query="what did the president say about ecology"
index.query(query)
```

Out[184]:
```
' The president said that the infrastructure plan will help withstand th
e devastating effects of the climate crisis and promote environmental ju
stice. He also said that they will build a national network of 500,000 e
lectric vehicle charging stations and begin to replace poisonous lead pi
pes so every American has clean water to drink.'
```

In [185…
```
query='what did he say about ecology'
index.query(query)
```

Out[185]:
```
' He said that his plan will promote environmental justice and withstand
the devastating effects of the climate crisis.'
```

In [186…
```
query='what did the president say about ketanji'
index.query_with_sources(query)
```

Out[186]:
```
{'question': 'what did the president say about ketanji',
 'answer': " The president mentioned Ketanji Brown Jackson, saying she i
s one of the nation's top legal minds and will continue Justice Breyer's
legacy of excellence.\n",
 'sources': '../input/state_of_the_union.txt'}
```

PDF files

(using pyPdf)

In [187…
```
# !pip install PyPDF
from langchain.document_loaders import PyPDFLoader
loader=PyPDFLoader("../input/2303.08774.pdf")
pages=loader.load_and_split()
pages[0]
```

Out[187]:
```
Document(page_content='GPT-4 Technical Report\nOpenAI\x03\nAbstract\nWe
report the development of GPT-4, a large-scale, multimodal model which c
an\naccept image and text inputs and produce text outputs. While less ca
pable than\nhumans in many real-world scenarios, GPT-4 exhibits human-le
vel performance\non various professional and academic benchmarks, includ
ing passing a simulated\nbar exam with a score around the top 10% of tes
t takers. GPT-4 is a Transformer-\nbased model pre-trained to predict th
e next token in a document. The post-training\nalignment process results
in improved performance on measures of factuality and\nadherence to desi
red behavior. A core component of this project was developing\ninfrastru
cture and optimization methods that behave predictably across a wide\nra
nge of scales. This allowed us to accurately predict some aspects of GPT
-4's\nperformance based on models trained with no more than 1/1,000th th
e compute of\nGPT-4.\n1 Introduction\nThis technical report presents GPT
-4, a large multimodal model capable of processing image and\ntext input
s and producing text outputs. Such models are an important area of study
as they have the\npotential to be used in a wide range of applications,
such as dialogue systems, text summarization,\nand machine translation.
As such, they have been the subject of substantial interest and progress
in\nrecent years [1–34].\nOne of the main goals of developing such model
s is to improve their ability to understand and generate\nnatural langua
ge text, particularly in more complex and nuanced scenarios. To test its
capabilities\nin such scenarios, GPT-4 was evaluated on a variety of exa
ms originally designed for humans. In\nthese evaluations it performs qui
te well and often outscores the vast majority of human test takers.\nFor
example, on a simulated bar exam, GPT-4 achieves a score that falls in t
he top 10% of test takers.\nThis contrasts with GPT-3.5, which scores in
the bottom 10%.\nOn a suite of traditional NLP benchmarks, GPT-4 outperf
orms both previous large language models\nand most state-of-the-art syst
ems (which often have benchmark-specific training or hand-engineering).\n
On the MMLU benchmark [ 35,36], an English-language suite of multiple-ch
oice questions covering\n57 subjects, GPT-4 not only outperforms existin
g models by a considerable margin in English, but\nalso demonstrates str
ong performance in other languages. On translated variants of MMLU, GPT-
4\nsurpasses the English-language state-of-the-art in 24 of 26 languages
considered. We discuss these\nmodel capability results, as well as model
safety improvements and results, in more detail in later\nsections.\nThi
s report also discusses a key challenge of the project, developing deep
learning infrastructure and\noptimization methods that behave predictabl
y across a wide range of scales. This allowed us to make\npredictions ab
out the expected performance of GPT-4 (based on small runs trained in si
milar ways)\nthat were tested against the final run to increase confidence
in our training.\nDespite its capabilities, GPT-4 has similar limitation
s to earlier GPT models [ 1,37,38]: it is not fully\nreliable (e.g. can
suffer from "hallucinations"), has a limited context window, and does no
t learn\n\x03Please cite this work as "OpenAI (2023)". Full authorship c
ontribution statements appear at the end of the\ndocument. Correspondenc
e regarding this technical report can be sent to gpt4-report@openai.coma
rXiv:2303.08774v3  [cs.CL]  27 Mar 2023', metadata={'source': '../input/
2303.08774.pdf', 'page': 0})
```

In [188…
```
from langchain.vectorstores import FAISS
from langchain.embeddings.openai import OpenAIEmbeddings
```

In [ ]:
```
faiss_index=FAISS.from_documents(pages,OpenAIEmbeddings())
docs=faiss_index.similarity_search("what are the deep learning strategies

for doc in docs:
    print(str(doc.metadata["page"]) + ":", doc.page_content)
```

Using Unstructured

In [ ]:
```
!pip install layoutparser # Install the base layoutparser library with
!pip install "layoutparser[layoutmodels]" # Install DL layout model toolk
!pip install "layoutparser[ocr]" # Install OCR toolkit
```

In [ ]:
```
# # Install packages
!pip install "unstructured[local-inference]"
!pip install "detectron2@git+https://github.com/facebookresearch/detectro
!pip install layoutparser[layoutmodels,tesseract]
```

In [ ]:
```
# # Install other dependencies
# # https://github.com/Unstructured-IO/unstructured/blob/main/docs/source
!brew install libmagic
!brew install poppler
!brew install tesseract
# If parsing xml / html documents:
!brew install libxml2
!brew install libxslt
```

In [190…
```
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     /Users/cherifbenham/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

Out[190]: True

In [191…
```
from langchain.document_loaders import UnstructuredFileLoader
```

In [192…
```
loader = UnstructuredFileLoader("../input/state_of_the_union.txt")
```

In [193…
```
docs = loader.load()
```

In [194…
```
docs[0].metadata
```

Out[194]: {'source': '../input/state_of_the_union.txt'}

In [195…
```
docs[0].page_content
```

Out[195]: 'Madam Speaker, Madam Vice President, our First Lady and Second Gentlema
n. Members of Congress and the Cabinet. Justices of the Supreme Court. M
y fellow Americans.\n\nLast year COVID-19 kept us apart. This year we ar
e finally together again.\n\nTonight, we meet as Democrats Republicans a
nd Independents. But most importantly as Americans.\n\nWith a duty to on
e another to the American people to the Constitution.\n\nAnd with an unw
avering resolve that freedom will always triumph over tyranny.\n\nSix da
ys ago, Russia's Vladimir Putin sought to shake the foundations of the f
ree world thinking he could make it bend to his menacing ways. But he ba
dly miscalculated.\n\nHe thought he could roll into Ukraine and the worl
d would roll over. Instead he met a wall of strength he never imagine
d.\n\nHe met the Ukrainian people.\n\nFrom President Zelenskyy to every
Ukrainian, their fearlessness, their courage, their determination, inspi
res the world.\n\nGroups of citizens blocking tanks with their bodies. E
veryone from students to retirees teachers turned soldiers defending the
ir homeland.\n\nIn this struggle as President Zelenskyy said in his spee
ch to the European Parliament "Light will win over darkness." The Ukrain
ian Ambassador to the United States is here tonight.\n\nLet each of us h
ere tonight in this Chamber send an unmistakable signal to Ukraine and t
o the world.\n\nPlease rise if you are able and show that, Yes, we the U
nited States of America stand with the Ukrainian people.\n\nThroughout o
ur history we've learned this lesson when dictators do not pay a price f
or their aggression they cause more chaos.\n\nThey keep moving.\n\nAnd t
he costs and the threats to America and the world keep rising.\n\nThat's
why the NATO Alliance was created to secure peace and stability in Europ
e after World War 2.\n\nThe United States is a member along with 29 othe
r nations.\n\nIt matters. American diplomacy matters. American resolve m
atters.\n\nPutin's latest attack on Ukraine was premeditated and unprovo
ked.\n\nHe rejected repeated efforts at diplomacy.\n\nHe thought the Wes
t and NATO wouldn't respond. And he thought he could divide us at home.
Putin was wrong. We were ready. Here is what we did.\n\nWe prepared ext
ensively and carefully.\n\nWe spent months building a coalition of other
freedom-loving nations from Europe and the Americas to Asia and Africa t
o confront Putin.\n\nI spent countless hours unifying our European allie
s. We shared with the world in advance what we knew Putin was planning a
nd precisely how he would try to falsely justify his aggression.\n\nWe c
ountered Russia's lies with truth.\n\nAnd now that he has acted the free
world is holding him accountable.\n\nAlong with twenty-seven members of
the European Union including France, Germany, Italy, as well as countrie
s like the United Kingdom, Canada, Japan, Korea, Australia, New Zealand,
and many others, even Switzerland.\n\nWe are inflicting pain on Russia a
nd supporting the people of Ukraine. Putin is now isolated from the worl
d more than ever.\n\nTogether with our allies –we are right now enforcin
g powerful economic sanctions.\n\nWe are cutting off Russia's largest ba
nks from the international financial system.\n\nPreventing Russia's cent
ral bank from defending the Russian Ruble making Putin's $630 Billion "w
ar fund" worthless.\n\nWe are choking off Russia's access to technology
that will sap its economic strength and weaken its military for years to
come.\n\nTonight I say to the Russian oligarchs and corrupt leaders who
have bilked billions of dollars off this violent regime no more.\n\nThe
U.S. Department of Justice is assembling a dedicated task force to go af
ter the crimes of Russian oligarchs.\n\nWe are joining with our European
allies to find and seize your yachts your luxury apartments your private
jets. We are coming for your ill-begotten gains.\n\nAnd tonight I am ann
ouncing that we will join our allies in closing off American air space t
o all Russian flights – further isolating Russia – and adding an additio
nal squeeze –on their economy. The Ruble has lost 30% of its value.\n\nT
he Russian stock market has lost 40% of its value and trading remains su
spended. Russia's economy is reeling and Putin alone is to blame.\n\nTo
gether with our allies we are providing support to the Ukrainians in th

eir fight for freedom. Military assistance. Economic assistance. Humanitarian assistance.\n\nWe are giving more than $1 Billion in direct assistance to Ukraine.\n\nAnd we will continue to aid the Ukrainian people as they defend their country and to help ease their suffering.\n\nLet me be clear, our forces are not engaged and will not engage in conflict with Russian forces in Ukraine.\n\nOur forces are not going to Europe to fight in Ukraine, but to defend our NATO Allies – in the event that Putin decides to keep moving west.\n\nFor that purpose we've mobilized American ground forces, air squadrons, and ship deployments to protect NATO countries including Poland, Romania, Latvia, Lithuania, and Estonia.\n\nAs I have made crystal clear the United States and our Allies will defend every inch of territory of NATO countries with the full force of our collective power.\n\nAnd we remain clear-eyed. The Ukrainians are fighting back with pure courage. But the next few days weeks, months, will be hard on them.\n\nPutin has unleashed violence and chaos. But while he may make gains on the battlefield – he will pay a continuing high price over the long run.\n\nAnd a proud Ukrainian people, who have known 30 years of independence, have repeatedly shown that they will not tolerate anyone who tries to take their country backwards.\n\nTo all Americans, I will be honest with you, as I've always promised. A Russian dictator, invading a foreign country, has costs around the world.\n\nAnd I'm taking robust action to make sure the pain of our sanctions is targeted at Russia's economy. And I will use every tool at our disposal to protect American businesses and consumers.\n\nTonight, I can announce that the United States has worked with 30 other countries to release 60 Million barrels of oil from reserves around the world.\n\nAmerica will lead that effort, releasing 30 Million barrels from our own Strategic Petroleum Reserve. And we stand ready to do more if necessary, unified with our allies.\n\nThese steps will help blunt gas prices here at home. And I know the news about what's happening can seem alarming.\n\nBut I want you to know that we are going to be okay.\n\nWhen the history of this era is written Putin's war on Ukraine will have left Russia weaker and the rest of the world stronger.\n\nWhile it shouldn't have taken something so terrible for people around the world to see what's at stake now everyone sees it clearly.\n\nWe see the unity among leaders of nations and a more unified Europe a more unified West. And we see unity among the people who are gathering in cities in large crowds around the world even in Russia to demonstrate their support for Ukraine.\n\nIn the battle between democracy and autocracy, democracies are rising to the moment, and the world is clearly choosing the side of peace and security.\n\nThis is a real test. It's going to take time. So let us continue to draw inspiration from the iron will of the Ukrainian people.\n\nTo our fellow Ukrainian Americans who forge a deep bond that connects our two nations we stand with you.\n\nPutin may circle Kyiv with tanks, but he will never gain the hearts and souls of the Ukrainian people.\n\nHe will never extinguish their love of freedom. He will never weaken the resolve of the free world.\n\nWe meet tonight in an America that has lived through two of the hardest years this nation has ever faced.\n\nThe pandemic has been punishing.\n\nAnd so many families are living paycheck to paycheck, struggling to keep up with the rising cost of food, gas, housing, and so much more.\n\nI understand.\n\nI remember when my Dad had to leave our home in Scranton, Pennsylvania to find work. I grew up in a family where if the price of food went up, you felt it.\n\nThat's why one of the first things I did as President was fight to pass the American Rescue Plan.\n\nBecause people were hurting. We needed to act, and we did.\n\nFew pieces of legislation have done more in a critical moment in our history to lift us out of crisis.\n\nIt fueled our efforts to vaccinate the nation and combat COVID-19. It delivered immediate economic relief for tens of millions of Americans.\n\nHelped put food on their table, keep a roof over their heads, and cut the cost of health in

surance.\n\nAnd as my Dad used to say, it gave people a little breathin g room.\n\nAnd unlike the $2 Trillion tax cut passed in the previous ad ministration that benefitted the top 1% of Americans, the American Resc ue Plan helped working people—and left no one behind.\n\nAnd it worked. It created jobs. Lots of jobs.\n\nIn fact—our economy created over 6.5 Million new jobs just last year, more jobs created in one year than eve r before in the history of America.\n\nOur economy grew at a rate of 5. 7% last year, the strongest growth in nearly 40 years, the first step i n bringing fundamental change to an economy that hasn't worked for the working people of this nation for too long.\n\nFor the past 40 years we were told that if we gave tax breaks to those at the very top, the bene fits would trickle down to everyone else.\n\nBut that trickle-down theo ry led to weaker economic growth, lower wages, bigger deficits, and the widest gap between those at the top and everyone else in nearly a centu ry.\n\nVice President Harris and I ran for office with a new economic v ision for America.\n\nInvest in America. Educate Americans. Grow the wo rkforce. Build the economy from the bottom up and the middle out, not f rom the top down.\n\nBecause we know that when the middle class grows, the poor have a ladder up and the wealthy do very well.\n\nAmerica used to have the best roads, bridges, and airports on Earth.\n\nNow our infr astructure is ranked 13th in the world.\n\nWe won't be able to compete for the jobs of the 21st Century if we don't fix that.\n\nThat's why it was so important to pass the Bipartisan Infrastructure Law—the most swe eping investment to rebuild America in history.\n\nThis was a bipartisa n effort, and I want to thank the members of both parties who worked to make it happen.\n\nWe're done talking about infrastructure weeks.\n\nW e're going to have an infrastructure decade.\n\nIt is going to transfor m America and put us on a path to win the economic competition of the 2 1st Century that we face with the rest of the world—particularly with C hina.\n\nAs I've told Xi Jinping, it is never a good bet to bet against the American people.\n\nWe'll create good jobs for millions of American s, modernizing roads, airports, ports, and waterways all across Americ a.\n\nAnd we'll do it all to withstand the devastating effects of the c limate crisis and promote environmental justice.\n\nWe'll build a natio nal network of 500,000 electric vehicle charging stations, begin to rep lace poisonous lead pipes—so every child—and every American—has clean w ater to drink at home and at school, provide affordable high-speed inte rnet for every American—urban, suburban, rural, and tribal communitie s.\n\n4,000 projects have already been announced.\n\nAnd tonight, I'm a nnouncing that this year we will start fixing over 65,000 miles of high way and 1,500 bridges in disrepair.\n\nWhen we use taxpayer dollars to rebuild America – we are going to Buy American: buy American products t o support American jobs.\n\nThe federal government spends about $600 Bi llion a year to keep the country safe and secure.\n\nThere's been a law on the books for almost a century to make sure taxpayers' dollars suppo rt American jobs and businesses.\n\nEvery Administration says they'll d o it, but we are actually doing it.\n\nWe will buy American to make sur e everything from the deck of an aircraft carrier to the steel on highw ay guardrails are made in America.\n\nBut to compete for the best jobs of the future, we also need to level the playing field with China and o ther competitors.\n\nThat's why it is so important to pass the Bipartis an Innovation Act sitting in Congress that will make record investments in emerging technologies and American manufacturing.\n\nLet me give you one example of why it's so important to pass it.\n\nIf you travel 20 mi les east of Columbus, Ohio, you'll find 1,000 empty acres of land.\n\nI t won't look like much, but if you stop and look closely, you'll see a "Field of dreams," the ground on which America's future will be buil t.\n\nThis is where Intel, the American company that helped build Silic on Valley, is going to build its $20 billion semiconductor "mega sit e".\n\nUp to eight state-of-the-art factories in one place. 10,000 new

good-paying jobs.\n\nSome of the most sophisticated manufacturing in th
e world to make computer chips the size of a fingertip that power the w
orld and our everyday lives.\n\nSmartphones. The Internet. Technology w
e have yet to invent.\n\nBut that's just the beginning.\n\nIntel's CEO,
Pat Gelsinger, who is here tonight, told me they are ready to increase
their investment from $20 billion to $100 billion.\n\nThat would be one
of the biggest investments in manufacturing in American history.\n\nAnd
all they're waiting for is for you to pass this bill.\n\nSo let's not w
ait any longer. Send it to my desk. I'll sign it.\n\nAnd we will really
take off.\n\nAnd Intel is not alone.\n\nThere's something happening in
America.\n\nJust look around and you'll see an amazing story.\n\nThe re
birth of the pride that comes from stamping products "Made In America."
The revitalization of American manufacturing.\n\nCompanies are choosing
to build new factories here, when just a few years ago, they would have
built them overseas.\n\nThat's what is happening. Ford is investing $11
billion to build electric vehicles, creating 11,000 jobs across the cou
ntry.\n\nGM is making the largest investment in its history—$7 billion
to build electric vehicles, creating 4,000 jobs in Michigan.\n\nAll tol
d, we created 369,000 new manufacturing jobs in America just last yea
r.\n\nPowered by people I've met like JoJo Burgess, from generations of
union steelworkers from Pittsburgh, who's here with us tonight.\n\nAs O
hio Senator Sherrod Brown says, "It's time to bury the label "Rust Bel
t."\n\nIt's time.\n\nBut with all the bright spots in our economy, reco
rd job growth and higher wages, too many families are struggling to kee
p up with the bills.\n\nInflation is robbing them of the gains they mig
ht otherwise feel.\n\nI get it. That's why my top priority is getting p
rices under control.\n\nLook, our economy roared back faster than most
predicted, but the pandemic meant that businesses had a hard time hirin
g enough workers to keep up production in their factories.\n\nThe pande
mic also disrupted global supply chains.\n\nWhen factories close, it ta
kes longer to make goods and get them from the warehouse to the store,
and prices go up.\n\nLook at cars.\n\nLast year, there weren't enough s
emiconductors to make all the cars that people wanted to buy.\n\nAnd gu
ess what, prices of automobiles went up.\n\nSo—we have a choice.\n\nOne
way to fight inflation is to drive down wages and make Americans poore
r.\n\nI have a better plan to fight inflation.\n\nLower your costs, not
your wages.\n\nMake more cars and semiconductors in America.\n\nMore in
frastructure and innovation in America.\n\nMore goods moving faster and
cheaper in America.\n\nMore jobs where you can earn a good living in Am
erica.\n\nAnd instead of relying on foreign supply chains, let's make i
t in America.\n\nEconomists call it "increasing the productive capacity
of our economy."\n\nI call it building a better America.\n\nMy plan to
fight inflation will lower your costs and lower the deficit.\n\n17 Nobe
l laureates in economics say my plan will ease long-term inflationary p
ressures. Top business leaders and most Americans support my plan. And
here's the plan:\n\nFirst — cut the cost of prescription drugs. Just lo
ok at insulin. One in ten Americans has diabetes. In Virginia, I met a
13-year-old boy named Joshua Davis.\n\nHe and his Dad both have Type 1
diabetes, which means they need insulin every day. Insulin costs about
$10 a vial to make.\n\nBut drug companies charge families like Joshua a
nd his Dad up to 30 times more. I spoke with Joshua's mom.\n\nImagine w
hat it's like to look at your child who needs insulin and have no idea
how you're going to pay for it.\n\nWhat it does to your dignity, your a
bility to look your child in the eye, to be the parent you expect to b
e.\n\nJoshua is here with us tonight. Yesterday was his birthday. Happy
birthday, buddy.\n\nFor Joshua, and for the 200,000 other young people
with Type 1 diabetes, let's cap the cost of insulin at $35 a month so e
veryone can afford it.\n\nDrug companies will still do very well. And w
hile we're at it let Medicare negotiate lower prices for prescription d
rugs, like the VA already does.\n\nLook, the American Rescue Plan is he

lping millions of families on Affordable Care Act plans save $2,400 a year on their health care premiums. Let's close the coverage gap and make those savings permanent.\n\nSecond – cut energy costs for families an average of $500 a year by combatting climate change.\n\nLet's provide investments and tax credits to weatherize your homes and businesses to be energy efficient and you get a tax credit; double America's clean energy production in solar, wind, and so much more; lower the price of electric vehicles, saving you another $80 a month because you'll never have to pay at the gas pump again.\n\nThird – cut the cost of child care. Many families pay up to $14,000 a year for child care per child.\n\nMiddle-class and working families shouldn't have to pay more than 7% of their income for care of young children.\n\nMy plan will cut the cost in half for most families and help parents, including millions of women, who left the workforce during the pandemic because they couldn't afford child care, to be able to get back to work.\n\nMy plan doesn't stop there. It also includes home and long-term care. More affordable housing. And Pre-K for every 3- and 4-year-old.\n\nAll of these will lower costs.\n\nAnd under my plan, nobody earning less than $400,000 a year will pay an additional penny in new taxes. Nobody.\n\nThe one thing all Americans agree on is that the tax system is not fair. We have to fix it.\n\nI'm not looking to punish anyone. But let's make sure corporations and the wealthiest Americans start paying their fair share.\n\nJust last year, 55 Fortune 500 corporations earned $40 billion in profits and paid zero dollars in federal income tax.\n\nThat's simply not fair. That's why I've proposed a 15% minimum tax rate for corporations.\n\nWe got more than 130 countries to agree on a global minimum tax rate so companies can't get out of paying their taxes at home by shipping jobs and factories overseas.\n\nThat's why I've proposed closing loopholes so the very wealthy don't pay a lower tax rate than a teacher or a firefighter.\n\nSo that's my plan. It will grow the economy and lower costs for families.\n\nSo what are we waiting for? Let's get this done. And while you're at it, confirm my nominees to the Federal Reserve, which plays a critical role in fighting inflation.\n\nMy plan will not only lower costs to give families a fair shot, it will lower the deficit.\n\nThe previous Administration not only ballooned the deficit with tax cuts for the very wealthy and corporations, it undermined the watchdogs whose job was to keep pandemic relief funds from being wasted.\n\nBut in my administration, the watchdogs have been welcomed back.\n\nWe're going after the criminals who stole billions in relief money meant for small businesses and millions of Americans.\n\nAnd tonight, I'm announcing that the Justice Department will name a chief prosecutor for pandemic fraud.\n\nBy the end of this year, the deficit will be down to less than half what it was before I took office.\n\nThe only president ever to cut the deficit by more than one trillion dollars in a single year.\n\nLowering your costs also means demanding more competition.\n\nI'm a capitalist, but capitalism without competition isn't capitalism.\n\nIt's exploitation—and it drives up prices.\n\nWhen corporations don't have to compete, their profits go up, your prices go up, and small businesses and family farmers and ranchers go under.\n\nWe see it happening with ocean carriers moving goods in and out of America.\n\nDuring the pandemic, these foreign-owned companies raised prices by as much as 1,000% and made record profits.\n\nTonight, I'm announcing a crackdown on these companies overcharging American businesses and consumers.\n\nAnd as Wall Street firms take over more nursing homes, quality in those homes has gone down and costs have gone up.\n\nThat ends on my watch.\n\nMedicare is going to set higher standards for nursing homes and make sure your loved ones get the care they deserve and expect.\n\nWe'll also cut costs and keep the economy going strong by giving workers a fair shot, provide more training and apprenticeships, hire them based on their skills not degrees.\n\nLet's pass the Paycheck Fairness Act and paid leave.\n\nRaise the

minimum wage to $15 an hour and extend the Child Tax Credit, so no one has to raise a family in poverty.\n\nLet’s increase Pell Grants and increase our historic support of HBCUs, and invest in what Jill—our First Lady who teaches full-time—calls America’s best-kept secret: community colleges.\n\nAnd let’s pass the PRO Act when a majority of workers want to form a union—they shouldn’t be stopped.\n\nWhen we invest in our workers, when we build the economy from the bottom up and the middle out together, we can do something we haven’t done in a long time: build a better America.\n\nFor more than two years, COVID-19 has impacted every decision in our lives and the life of the nation.\n\nAnd I know you’re tired, frustrated, and exhausted.\n\nBut I also know this.\n\nBecause of the progress we’ve made, because of your resilience and the tools we have, tonight I can say we are moving forward safely, back to more normal routines.\n\nWe’ve reached a new moment in the fight against COVID-19, with severe cases down to a level not seen since last July.\n\nJust a few days ago, the Centers for Disease Control and Prevention—the CDC—issued new mask guidelines.\n\nUnder these new guidelines, most Americans in most of the country can now be mask free.\n\nAnd based on the projections, more of the country will reach that point across the next couple of weeks.\n\nThanks to the progress we have made this past year, COVID-19 need no longer control our lives.\n\nI know some are talking about “living with COVID-19”. Tonight – I say that we will never just accept living with COVID-19.\n\nWe will continue to combat the virus as we do other diseases. And because this is a virus that mutates and spreads, we will stay on guard.\n\nHere are four common sense steps as we move forward safely.\n\nFirst, stay protected with vaccines and treatments. We know how incredibly effective vaccines are. If you’re vaccinated and boosted you have the highest degree of protection.\n\nWe will never give up on vaccinating more Americans. Now, I know parents with kids under 5 are eager to see a vaccine authorized for their children.\n\nThe scientists are working hard to get that done and we’ll be ready with plenty of vaccines when they do.\n\nWe’re also ready with anti-viral treatments. If you get COVID-19, the Pfizer pill reduces your chances of ending up in the hospital by 90%.\n\nWe’ve ordered more of these pills than anyone in the world. And Pfizer is working overtime to get us 1 Million pills this month and more than double that next month.\n\nAnd we’re launching the “Test to Treat” initiative so people can get tested at a pharmacy, and if they’re positive, receive antiviral pills on the spot at no cost.\n\nIf you’re immunocompromised or have some other vulnerability, we have treatments and free high-quality masks.\n\nWe’re leaving no one behind or ignoring anyone’s needs as we move forward.\n\nAnd on testing, we have made hundreds of millions of tests available for you to order for free.\n\nEven if you already ordered free tests tonight, I am announcing that you can order more from covidtests.gov starting next week.\n\nSecond – we must prepare for new variants. Over the past year, we’ve gotten much better at detecting new variants.\n\nIf necessary, we’ll be able to deploy new vaccines within 100 days instead of many more months or years.\n\nAnd, if Congress provides the funds we need, we’ll have new stockpiles of tests, masks, and pills ready if needed.\n\nI cannot promise a new variant won’t come. But I can promise you we’ll do everything within our power to be ready if it does.\n\nThird – we can end the shutdown of schools and businesses. We have the tools we need.\n\nIt’s time for Americans to get back to work and fill our great downtowns again. People working from home can feel safe to begin to return to the office.\n\nWe’re doing that here in the federal government. The vast majority of federal workers will once again work in person.\n\nOur schools are open. Let’s keep it that way. Our kids need to be in school.\n\nAnd with 75% of adult Americans fully vaccinated and hospitalizations down by 77%, most Americans can remove their masks, return to work, stay in the classroom, and move forward safely.\n\nWe achieved this

because we provided free vaccines, treatments, tests, and masks.\n\nOf course, continuing this costs money.\n\nI will soon send Congress a request.\n\nThe vast majority of Americans have used these tools and may want to again, so I expect Congress to pass it quickly.\n\nFourth, we will continue vaccinating the world.\n\nWe've sent 475 Million vaccine doses to 112 countries, more than any other nation.\n\nAnd we won't stop.\n\nWe have lost so much to COVID-19. Time with one another. And worst of all, so much loss of life.\n\nLet's use this moment to reset. Let's stop looking at COVID-19 as a partisan dividing line and see it for what it is: A God-awful disease.\n\nLet's stop seeing each other as enemies, and start seeing each other for who we really are: Fellow Americans.\n\nWe can't change how divided we've been. But we can change how we move forward—on COVID-19 and other issues we must face together.\n\nI recently visited the New York City Police Department days after the funerals of Officer Wilbert Mora and his partner, Officer Jason Rivera.\n\nThey were responding to a 9-1-1 call when a man shot and killed them with a stolen gun.\n\nOfficer Mora was 27 years old.\n\nOfficer Rivera was 22.\n\nBoth Dominican Americans who'd grown up on the same streets they later chose to patrol as police officers.\n\nI spoke with their families and told them that we are forever in debt for their sacrifice, and we will carry on their mission to restore the trust and safety every community deserves.\n\nI've worked on these issues a long time.\n\nI know what works: Investing in crime preventionand community police officers who'll walk the beat, who'll know the neighborhood, and who can restore trust and safety.\n\nSo let's not abandon our streets. Or choose between safety and equal justice.\n\nLet's come together to protect our communities, restore trust, and hold law enforcement accountable.\n\nThat's why the Justice Department required body cameras, banned chokeholds, and restricted no-knock warrants for its officers.\n\nThat's why the American Rescue Plan provided $350 Billion that cities, states, and counties can use to hire more police and invest in proven strategies like community violence interruption—trusted messengers breaking the cycle of violence and trauma and giving young people hope.\n\nWe should all agree: The answer is not to Defund the police. The answer is to FUND the police with the resources and training they need to protect our communities.\n\nI ask Democrats and Republicans alike: Pass my budget and keep our neighborhoods safe.\n\nAnd I will keep doing everything in my power to crack down on gun trafficking and ghost guns you can buy online and make at home—they have no serial numbers and can't be traced.\n\nAnd I ask Congress to pass proven measures to reduce gun violence. Pass universal background checks. Why should anyone on a terrorist list be able to purchase a weapon?\n\nBan assault weapons and high-capacity magazines.\n\nRepeal the liability shield that makes gun manufacturers the only industry in America that can't be sued.\n\nThese laws don't infringe on the Second Amendment. They save lives.\n\nThe most fundamental right in America is the right to vote – and to have it counted. And it's under assault.\n\nIn state after state, new laws have been passed, not only to suppress the vote, but to subvert entire elections.\n\nWe cannot let this happen.\n\nTonight. I call on the Senate to: Pass the Freedom to Vote Act. Pass the John Lewis Voting Rights Act. And while you're at it, pass the Disclose Act so Americans can know who is funding our elections.\n\nTonight, I'd like to honor someone who has dedicated his life to serve this country: Justice Stephen Breyer—an Army veteran, Constitutional scholar, and retiring Justice of the United States Supreme Court. Justice Breyer, thank you for your service.\n\nOne of the most serious constitutional responsibilities a President has is nominating someone to serve on the United States Supreme Court.\n\nAnd I did that 4 days ago, when I nominated Circuit Court of Appeals Judge Ketanji Brown Jackson. One of our nation's top legal minds, who will continue Justice Breyer's legacy of excellence.\n\nA former top litigator in private practice. A

former federal public defender. And from a family of public school educ
ators and police officers. A consensus builder. Since she’s been nomina
ted, she’s received a broad range of support—from the Fraternal Order o
f Police to former judges appointed by Democrats and Republicans.\n\nAn
d if we are to advance liberty and justice, we need to secure the Borde
r and fix the immigration system.\n\nWe can do both. At our border, w
e’ve installed new technology like cutting-edge scanners to better dete
ct drug smuggling.\n\nWe’ve set up joint patrols with Mexico and Guatem
ala to catch more human traffickers.\n\nWe’re putting in place dedicate
d immigration judges so families fleeing persecution and violence can h
ave their cases heard faster.\n\nWe’re securing commitments and support
ing partners in South and Central America to host more refugees and sec
ure their own borders.\n\nWe can do all this while keeping lit the torc
h of liberty that has led generations of immigrants to this land—my for
efathers and so many of yours.\n\nProvide a pathway to citizenship for
Dreamers, those on temporary status, farm workers, and essential worker
s.\n\nRevise our laws so businesses have the workers they need and fami
lies don’t wait decades to reunite.\n\nIt’s not only the right thing to
do—it’s the economically smart thing to do.\n\nThat’s why immigration r
eform is supported by everyone from labor unions to religious leaders t
o the U.S. Chamber of Commerce.\n\nLet’s get it done once and for al
l.\n\nAdvancing liberty and justice also requires protecting the rights
of women.\n\nThe constitutional right affirmed in Roe v. Wade—standing
precedent for half a century—is under attack as never before.\n\nIf we
want to go forward—not backward—we must protect access to health care.
Preserve a woman’s right to choose. And let’s continue to advance mater
nal health care in America.\n\nAnd for our LGBTQ+ Americans, let’s fina
lly get the bipartisan Equality Act to my desk. The onslaught of state
laws targeting transgender Americans and their families is wrong.\n\nAs
I said last year, especially to our younger transgender Americans, I wi
ll always have your back as your President, so you can be yourself and
reach your God-given potential.\n\nWhile it often appears that we never
agree, that isn’t true. I signed 80 bipartisan bills into law last yea
r. From preventing government shutdowns to protecting Asian-Americans f
rom still-too-common hate crimes to reforming military justice.\n\nAnd
soon, we’ll strengthen the Violence Against Women Act that I first wrot
e three decades ago. It is important for us to show the nation that we
can come together and do big things.\n\nSo tonight I’m offering a Unity
Agenda for the Nation. Four big things we can do together.\n\nFirst, be
at the opioid epidemic.\n\nThere is so much we can do. Increase funding
for prevention, treatment, harm reduction, and recovery.\n\nGet rid of
outdated rules that stop doctors from prescribing treatments. And stop
the flow of illicit drugs by working with state and local law enforceme
nt to go after traffickers.\n\nIf you’re suffering from addiction, know
you are not alone. I believe in recovery, and I celebrate the 23 millio
n Americans in recovery.\n\nSecond, let’s take on mental health. Especi
ally among our children, whose lives and education have been turned ups
ide down.\n\nThe American Rescue Plan gave schools money to hire teache
rs and help students make up for lost learning.\n\nI urge every parent
to make sure your school does just that. And we can all play a part—sig
n up to be a tutor or a mentor.\n\nChildren were also struggling before
the pandemic. Bullying, violence, trauma, and the harms of social medi
a.\n\nAs Frances Haugen, who is here with us tonight, has shown, we must
hold social media platforms accountable for the national experiment the
y’re conducting on our children for profit.\n\nIt’s time to strengthen p
rivacy protections, ban targeted advertising to children, demand tech co
mpanies stop collecting personal data on our children.\n\nAnd let’s get
all Americans the mental health services they need. More people they can
turn to for help, and full parity between physical and mental health car
e.\n\nThird, support our veterans.\n\nVeterans are the best of us.\n\n

I’ve always believed that we have a sacred obligation to equip all those we send to war and care for them and their families when they come hom e.\n\nMy administration is providing assistance with job training and ho using, and now helping lower-income veterans get VA care debt-free.\n\nO ur troops in Iraq and Afghanistan faced many dangers.\n\nOne was station ed at bases and breathing in toxic smoke from “burn pits” that incinerat ed wastes of war—medical and hazard material, jet fuel, and more.\n\nWhe n they came home, many of the world’s fittest and best trained warriors were never the same.\n\nHeadaches. Numbness. Dizziness.\n\nA cancer that would put them in a flag-draped coffin.\n\nI know.\n\nOne of those soldi ers was my son Major Beau Biden.\n\nWe don’t know for sure if a burn pit was the cause of his brain cancer, or the diseases of so many of our tro ops.\n\nBut I’m committed to finding out everything we can.\n\nCommitted to military families like Danielle Robinson from Ohio.\n\nThe widow of S ergeant First Class Heath Robinson.\n\nHe was born a soldier. Army Natio nal Guard. Combat medic in Kosovo and Iraq.\n\nStationed near Baghdad, j ust yards from burn pits the size of football fields.\n\nHeath’s widow D anielle is here with us tonight. They loved going to Ohio State football games. He loved building Legos with their daughter.\n\nBut cancer from p rolonged exposure to burn pits ravaged Heath’s lungs and body.\n\nDaniel le says Heath was a fighter to the very end.\n\nHe didn’t know how to st op fighting, and neither did she.\n\nThrough her pain she found purpose to demand we do better.\n\nTonight, Danielle—we are.\n\nThe VA is pionee ring new ways of linking toxic exposures to diseases, already helping mo re veterans get benefits.\n\nAnd tonight, I’m announcing we’re expanding eligibility to veterans suffering from nine respiratory cancers.\n\nI’m also calling on Congress: pass a law to make sure veterans devastated by toxic exposures in Iraq and Afghanistan finally get the benefits and com prehensive health care they deserve.\n\nAnd fourth, let’s end cancer as we know it.\n\nThis is personal to me and Jill, to Kamala, and to so man y of you.\n\nCancer is the #2 cause of death in America—second only to h eart disease.\n\nLast month, I announced our plan to supercharge the Can cer Moonshot that President Obama asked me to lead six years ago.\n\nOur goal is to cut the cancer death rate by at least 50% over the next 25 ye ars, turn more cancers from death sentences into treatable diseases.\n\n More support for patients and families.\n\nTo get there, I call on Congr ess to fund ARPA-H, the Advanced Research Projects Agency for Health.\n \nIt’s based on DARPA—the Defense Department project that led to the Int ernet, GPS, and so much more.\n\nARPA-H will have a singular purpose—to drive breakthroughs in cancer, Alzheimer’s, diabetes, and more.\n\nA uni ty agenda for the nation.\n\nWe can do this.\n\nMy fellow Americans—toni ght , we have gathered in a sacred space—the citadel of our democracy.\n \nIn this Capitol, generation after generation, Americans have debated g reat questions amid great strife, and have done great things.\n\nWe have fought for freedom, expanded liberty, defeated totalitarianism and terro r.\n\nAnd built the strongest, freest, and most prosperous nation the wo rld has ever known.\n\nNow is the hour.\n\nOur moment of responsibilit y.\n\nOur test of resolve and conscience, of history itself.\n\nIt is in this moment that our character is formed. Our purpose is found. Our futu re is forged.\n\nWell I know this nation.\n\nWe will meet the test.\n\nT o protect freedom and liberty, to expand fairness and opportunity.\n\nWe will save democracy.\n\nAs hard as these times have been, I am more opti mistic about America today than I have been my whole life.\n\nBecause I see the future that is within our grasp.\n\nBecause I know there is simp ly nothing beyond our capacity.\n\nWe are the only nation on Earth that has always turned every crisis we have faced into an opportunity.\n\nThe only nation that can be defined by a single word: possibilities.\n\nSo o n this night, in our 245th year as a nation, I have come to report on th e State of the Union.\n\nAnd my report is this: the State of the Union i s strong—because you, the American people, are strong.\n\nWe are stronge

```
r today than we were a year ago.\n\nAnd we will be stronger a year from
now than we are today.\n\nNow is our moment to meet and overcome the cha
llenges of our time.\n\nAnd we will, as one people.\n\nOne America.\n\nT
he United States of America.\n\nMay God bless you all. May God protect o
ur troops.'
```

Under the hood, Unstructured creates different "elements" for different chunks of text. By default we combine those together, but you can easily keep that separation by specifying mode="elements".

In [196…
```
loader = UnstructuredFileLoader("../input/state_of_the_union.txt", mode="
docs=loader.load()
docs[:5]
```

Out[196]:
```
[Document(page_content='Madam Speaker, Madam Vice President, our First L
ady and Second Gentleman. Members of Congress and the Cabinet. Justices
of the Supreme Court. My fellow Americans.', metadata={'source': '../inp
ut/state_of_the_union.txt', 'filename': '../input/state_of_the_union.tx
t', 'category': 'NarrativeText'}),
 Document(page_content='Last year COVID-19 kept us apart. This year we a
re finally together again.', metadata={'source': '../input/state_of_the_
union.txt', 'filename': '../input/state_of_the_union.txt', 'category':
'NarrativeText'}),
 Document(page_content='Tonight, we meet as Democrats Republicans and In
dependents. But most importantly as Americans.', metadata={'source':
'../input/state_of_the_union.txt', 'filename': '../input/state_of_the_un
ion.txt', 'category': 'NarrativeText'}),
 Document(page_content='With a duty to one another to the American peopl
e to the Constitution.', metadata={'source': '../input/state_of_the_unio
n.txt', 'filename': '../input/state_of_the_union.txt', 'category': 'Unca
tegorizedText'}),
 Document(page_content='And with an unwavering resolve that freedom will
always triumph over tyranny.', metadata={'source': '../input/state_of_th
e_union.txt', 'filename': '../input/state_of_the_union.txt', 'category':
'NarrativeText'})]
```

Define a Partitioning Strategy

Unstructured document loader allow users to pass in a strategy parameter that lets unstructured know how to partitioning the document. Currently supported strategies are "hi_res" (the default) and "fast". Hi res partitioning strategies are more accurate, but take longer to process. Fast strategies partition the document more quickly, but trade-off accuracy. Not all document types have separate hi res and fast partitioning strategies. For those document types, the strategy kwarg is ignored. In some cases, the high res strategy will fallback to fast if there is a dependency missing (i.e. a model for document partitioning). You can see how to apply a strategy to an UnstructuredFileLoader below.

Another PDF Example

In [ ]:
```
!wget  https://raw.githubusercontent.com/Unstructured-IO/unstructured/mai
```

In [197…
```
loader = UnstructuredFileLoader("../input/layout-parser-paper.pdf", mode=
```

In [198…
```
docs = loader.load()
```

In [199…
```
docs[:10]
```

Out[199]:
```
[Document(page_content='LayoutParser: A Unified Toolkit for Deep Learnin
g Based Document Image Analysis', metadata={'source': '../input/layout-p
arser-paper.pdf', 'filename': '../input/layout-parser-paper.pdf', 'page_
number': 1, 'category': 'Title'}),
 Document(page_content='Zejiang Shen1 ((cid:0)), Ruochen Zhang2, Melissa
Dell3, Benjamin Charles Germain Lee4, Jacob Carlson3, and Weining Li5',
metadata={'source': '../input/layout-parser-paper.pdf', 'filename': '../
input/layout-parser-paper.pdf', 'page_number': 1, 'category': 'Narrative
Text'}),
 Document(page_content='Allen Institute for AI shannons@allenai.org', me
tadata={'source': '../input/layout-parser-paper.pdf', 'filename': '../in
put/layout-parser-paper.pdf', 'page_number': 1, 'category': 'ListIte
m'}),
 Document(page_content='Brown University ruochen zhang@brown.edu', metad
ata={'source': '../input/layout-parser-paper.pdf', 'filename': '../inpu
t/layout-parser-paper.pdf', 'page_number': 1, 'category': 'ListItem'}),
 Document(page_content='Harvard University {melissadell,jacob', metadata
={'source': '../input/layout-parser-paper.pdf', 'filename': '../input/la
yout-parser-paper.pdf', 'page_number': 1, 'category': 'ListItem'}),
 Document(page_content='University of Washington bcgl@cs.washington.ed
u', metadata={'source': '../input/layout-parser-paper.pdf', 'filename':
'../input/layout-parser-paper.pdf', 'page_number': 1, 'category': 'ListI
tem'}),
 Document(page_content='University of Waterloo w', metadata={'source':
'../input/layout-parser-paper.pdf', 'filename': '../input/layout-parser-
paper.pdf', 'page_number': 1, 'category': 'ListItem'}),
 Document(page_content='li@uwaterloo.ca', metadata={'source': '../input/
layout-parser-paper.pdf', 'filename': '../input/layout-parser-paper.pd
f', 'page_number': 1, 'category': 'ListItem'}),
 Document(page_content='Abstract. Recent advances in document image anal
ysis (DIA) have been primarily driven by the application of neural netwo
rks. Ideally, research outcomes could be easily deployed in production a
nd extended for further investigation. However, various factors like loo
sely organized codebases and sophisticated model configurations complica
te the easy reuse of im- portant innovations by a wide audience. Though
there have been on-going efforts to improve reusability and simplify dee
p learning (DL) model development in disciplines like natural language p
rocessing and computer vision, none of them are optimized for challenges
in the domain of DIA. This represents a major gap in the existing toolki
t, as DIA is central to academic research across a wide range of discipl
ines in the social sciences and humanities. This paper introduces Layout
Parser, an open-source library for streamlining the usage of DL in DIA r
esearch and applica- tions. The core LayoutParser library comes with a s
et of simple and intuitive interfaces for applying and customizing DL mo
dels for layout de- tection, character recognition, and many other docum
ent processing tasks. To promote extensibility, LayoutParser also incorp
orates a community platform for sharing both pre-trained models and full
document digiti- zation pipelines. We demonstrate that LayoutParser is h
elpful for both lightweight and large-scale digitization pipelines in re
al-word use cases. The library is publicly available at https://layout-p
arser.github.io.', metadata={'source': '../input/layout-parser-paper.pd
f', 'filename': '../input/layout-parser-paper.pdf', 'page_number': 1, 'c
ategory': 'NarrativeText'}),
 Document(page_content='Keywords: Document Image Analysis Deep Learning
Layout · Character Recognition · Open Source library · Toolkit.', metada
ta={'source': '../input/layout-parser-paper.pdf', 'filename': '../input/
layout-parser-paper.pdf', 'page_number': 1, 'category': 'NarrativeTex
t'})]
```

URL Loader

This covers how to load HTML documents from a list of URLs into a document format that we can use downstream.

In [ ]:
```
from langchain.document_loaders import UnstructuredURLLoader
```

In [200…
```
urls = [
    "https://www.understandingwar.org/backgrounder/russian-offensive-camp
    "https://www.understandingwar.org/backgrounder/russian-offensive-camp
]
```

In [201…
```
loader = UnstructuredURLLoader(urls=urls)
```

In [202…
```
docs=loader.load()
```

In [203…
```
docs
```

Out[203]: [Document(page_content='Skip to main content\n\nSearch form\n\nHome\n\nWho We Are\n\nResearch\n\nPublications\n\nGet Involved\n\nPlanned Giving\n\nDonate\n\nRussian Offensive Campaign Assessment, February 8, 2023\n\nFeb 8, 2023 - Press ISW\n\nDownload the PDF\n\nKarolina Hird, Riley Bailey, George Barros, Layne Philipson, Nicole Wolkov, and Mason Clark\n\nFebruary 8, 8:30pm ET\n\nClick\xa0here\xa0to see ISW's interactive map of the Russian invasion of Ukraine. This map is updated daily alongside the static maps present in this report.\n\n[1]\xa0Geolocated combat footage has confirmed Russian gains in the Dvorichne area northwest of Svatove.\n\n[2]\xa0Russian military command additionally appears to have fully committed elements of several conventional divisions to decisive offensive operations along the Svatove-Kreminna line, as ISW previously reported.\n\n[3]\xa0Elements of several regiments of the 144\n\n[4]\n\nThe commitment of significant elements of at least three major Russian divisions to offensive operations in this sector indicates the Russian offensive has begun, even if Ukrainian forces are so far preventing Russian forces from securing significant gains.\xa0The Russian offensive likely has not yet reached its full tempo; Russian command has not yet committed elements of the 2nd\xa0Motorized Rifle Division (1st\xa0Guards Tank Army, Western Military District), which deployed to Luhansk Oblast in January after deploying to Belarus.[5]\xa0Russian forces are gradually beginning an offensive, but its success is not inherent or predetermined. While Russian forces in Luhansk Oblast now have the initiative (in that Russian forces are setting the terms of battle, ending the period of Ukrainian initiative from August 2022), the full commitment of these forces could lead to their eventual culmination along the Svatove-Kreminna line without achieving their objectives of capturing all of Luhansk and Donetsk oblasts. That culmination would likely provide a window of opportunity for Ukrainian forces to exploit with their own counteroffensive.[6]\n\nDonetsk People's Republic (DNR) People's Militia command reportedly assumed control over a Russian artillery battalion, likely in support of an effort to strengthen degraded DNR forces ahead of an imminent Russian offensive.\xa0A Russian source published a video appeal from mobilized personnel of the 640th\xa0howitzer battalion from Saratov Oblast on February 8 in which they stated that Russian military officials sent them to join DNR units and that DNR commanders are now trying to transfer them to infantry assault units.[7]\xa0ISW has not previously observed Russian personnel subordinated to a DNR formation and this claim, if true, would suggest that Russian forces may be reinforcing degraded DNR formations with mobilized personnel from Russia itself because DNR formations are unable to replenish losses themselves. The reported subordination of Russian military personnel to DNR formations may portend a Russian effort to prepare DNR formations for an expanded role in their zone of responsibility along the western outskirts of Donetsk City, and the transfer of remaining conventional Russian forces from this area to the Bakhmut area and Luhansk Oblast, where Russian forces are conducting an increased pace of offensive operations.\n\n[8]\xa0Russian forces would likely need to temporarily remove these irregular forces from frontline positions to integrate them into new Russian formations, a prospect that would not be operationally sound ahead of increased Russan offensive operations in Ukraine. Russian officials therefore may be attempting to gradually integrate these irregular formations through subordinating mobilized personnel under them without disrupting the command structures and existing personnel operating at front line positions. The mobilized personnel of the 640\n\n[9]\xa0The Russian MoD will likely struggle to correct the poor effectiveness of DNR/LNR forces through the rapid integration of Russian personnel.\n\nRussian officials continue to propose measures to prepare Russia's military industry for a protracted war in Ukraine while also likely setting further conditions for sanctions evasion.\xa0Russian Prime Minister Mikhail Mishustin stated on February 8 that the Russian government will subsi

dize investment projects for the modernization of enterprises operating in the interests of the Russian military and will allocate significant funds for manufacturing new military equipment.[10]\xa0Mishustin also stated that the Russian government would extend benefits to Russian entrepreneurs who support the Russian military, including extended payment periods on rented federal property.[11]\xa0The Kremlin likely intends these measures to augment its overarching effort to gradually prepare Russia's military industry for a protracted war in Ukraine while avoiding a wider economic mobilization that would create further domestic economic disruptions and corresponding discontent.[12]\n\nRussian officials also likely proposed these measures in coordination with a recent decree excluding Russian officials from requirements to list income declarations and proposals to repeal federal procurement procedures. The Kremlin may be creating a system of subsidies and benefits designed to have little oversight or accounting. This lack of oversight and accounting would likely allow Russian firms to better evade international sanctions regimes targeting Russia's military industry.[13]\xa0The United Kingdom announced a new list of sanctioned entities on February 8 focused on Russia's military industry.[14]\xa0ISW previously reported that 82% of Iranian-made drones downed in Ukraine had chips, semiconductors, and other components from the United States, suggesting that Russia and Iran are likely exploiting loopholes to transfer Western-produced arms components to Russia via proxy actors.[15]\xa0The Kremlin's effort to prepare the Russian military industry for a protracted war in Ukraine in part relies on the ability of Russian military industry to have consistent access to multiple secure supply chains of key foreign components that it otherwise cannot produce.\n\nKey Takeaways\n\nRussian forces have regained the initiative in Ukraine and have begun their next major offensive in Luhansk Oblast.\n\nThe commitment of significant elements of at least three major Russian divisions to offensive operations in this sector indicates the Russian offensive has begun, even if Ukrainian forces are so far preventing Russian forces from securing significant gains.\n\nDonetsk People's Republic (DNR) People's Militia command reportedly assumed control over a Russian artillery battalion, likely in support of an effort to strengthen degraded DNR forces ahead of an imminent Russian offensive.\n\nThe reported subordination of Russian mobilized personnel to DNR formations could also suggest that Russian military command may be continuing efforts to integrate ad hoc DNR and Luhansk People's Republic (LNR) formations into the Russian Armed Forces, but will likely face significant difficulties.\n\nRussian officials continue to propose measures to prepare Russia's military industry for a protracted war in Ukraine while also likely setting further conditions for sanctions evasion.\n\nRussian forces conducted ground attacks around Bakhmut and continued making tactical advances.\n\nRussian forces continued offensive actions northwest of Svatove and intensified offensive operations near Kreminna.\n\nRussian forces conducted limited ground attacks in the Avdiivka-Donetsk City area and western Donetsk Oblast.\n\nRussian and Ukrainian forces reportedly continue small-scale skirmishes and reconnaissance activity in the Dnipro River delta and on the Kinburn Spit.\n\nThe Wagner Group is reportedly resorting to more coercive tactics in its prison recruitment campaign, possibly in response to the campaign's declining effectiveness.\n\nWe do not report in detail on Russian war crimes because those activities are well-covered in Western media and do not directly affect the military operations we are assessing and forecasting. We will continue to evaluate and report on the effects of these criminal activities on the Ukrainian military and population and specifically on combat in Ukrainian urban areas. We utterly condemn these Russian violations of the laws of armed conflict, Geneva Conventions, and humanity even though we do not describe them in these reports.\n\nUkrainian Counteroffensives—Eastern Ukraine\n\nRussian Main Effort—Eastern Ukraine

(comprised of two subordinate main efforts);\n\nRussia Subordinate Main Effort #1—Capture the remainder of Luhansk Oblast and push westward into eastern Kharkiv Oblast and encircle northern Donetsk Oblast\n\nRussian Subordinate Main Effort #2—Capture the entirety of Donetsk Oblast\n\nRussian Supporting Effort—Southern Axis\n\nRussian Mobilization and Force Generation Efforts\n\nActivities in Russian-occupied Areas\n\nRussian Main Effort—Eastern Ukraine\n\nRussian Subordinate Main Effort #1- Luhansk Oblast (Russian objective: Capture the remainder of Luhansk Oblast and continue offensive operations into eastern Kharkiv Oblast and\xa0\xa0northern Donetsk Oblast)\n\nISW continues to assess the current Russian most likely course of action (MLCOA) is an imminent offensive effort in Luhansk Oblast and is therefore adjusting the structure of the daily campaign assessments. We will no longer include the Eastern Kharkiv and Western Luhansk Oblast area as part of Ukrainian counteroffensives and will assess this area as a subordinate part of the Russian main effort in Eastern Ukraine. The assessment of Luhansk Oblast as part of the Russian main effort does not preclude the possibility of continued Ukrainian counteroffensive actions here or anywhere else in theater in the future. ISW will report out on Ukrainian counteroffensive efforts as they occur.\n\nRussian forces continued offensive actions northwest of Svatove on February 8. Kharkiv Oblast Head Oleh Synehubov reported on February 8 that Russian forces are increasing their presence northwest of Svatove in the Kupyansk and Dvorichna areas.[16]\xa0A former Luhansk People‘s Republic (LNR) deputy claimed that fierce fighting is ongoing 7km from the Kupyansk area, likely referring to areas near Synkivka, which Russian sources claimed Russian forces captured on February 6.[17]\xa0The Ukrainian General Staff reported that Russian forces conducted a limited ground attack near Novoselivske, about 15km northwest of Svatove.[18]\xa0\xa0Former Russian militant commander and nationalist milblogger Igor Girkin denied that Russian forces have made any significant territorial gains in Kharkiv Oblast, particularly in the Kupyansk direction, as of February 8.[19]\n\n[20]\xa0The Ukrainian General Staff reported that Russian forces attacked near Chervonopopivka (5km north of Kreminna).\n\n[21]\xa0Several Russian milbloggers circulated unconfirmed footage of unspecified Central and Western Military District elements which crossed the Zherebets River running north to south in western Luhansk Oblast, roughly parallel to the Svatove-Kreminna line) and captured Ukrainian positions in an unspecified location around February 6.\n\n[22]\xa0Russian sources also reported that elements of the 3\n\n[23]\xa0A prominent Russian milblogger posted footage of the 59\n\n[24]\n\nRussian forces continued offensive operations south of Kreminna on February 8. The Ukrainian General Staff reported Russian troops attacked near Shepilove (7km south of Kreminna) and Bilohorivka (10km south of Kreminna).[25]\xa0Chechen Head Ramzan Kadyrov claimed that elements of the Chechen ”Akhmat” special forces and 2nd\xa0Brigade of the Luhansk People’s Republic 2nd\xa0Army Corps captured Ukrainian positions near Berestove, 30km south of Kreminna.[26]\xa0Russian forces appear to be pushing northeast of the Bakhmut area towards Siversk (17km southwest of Kreminna) to provide a supporting line of advance to the likely main Russian push directly westward toward Kreminna.\n\nRussian Subordinate Main Effort #2—Donetsk Oblast\xa0(Russian objective: Capture the entirety of Donetsk Oblast, the claimed territory of Russia’s proxies in Donbas)\n\nRussian forces conducted ground attacks around Bakhmut and continued making tactical advances on February 8. Geolocated footage posted between February 4 and 8 confirms that Russian forces have made marginal advances north of Bakhmut near Krasna Hora and Zaliznyanske (10km north of Bakhmut), in the Stupky area of northern Bakhmut, and southwest of Bakhmut near Ivanivske.[27]\xa0Russian forces are visually confirmed to be within 2.5 km of the E40 Bakhmut-Slovyansk highway.[28]\xa0The Ukrainian General Staff also reported that Ukrainian troops repelled Russian attacks on Ba

khmut itself; northeast of Bakhmut near Verkhnokamyanske (30km northeast), Fedorivka (15km northeast), Spirne (27km northeast), and Vymika (20 km northeast); north of Bakhmut near Paraskoviivka (5km north) and Krasna Hora (4km north); northwest of Bakhmut near Orikhovo-Vasylivka (12km northwest) and Dubovo-Vasylivka (7km northwest); and west of Bakhmut near Ivanivske (5km west) and Chasiv Yar (10km west).[29]\xa0\xa0The Ukrainian General Staff's report that Russian forces are attacking towards Orikhovo-Vasylivka and Dubovo-Vasylivka is consistent with geolocated combat footage and indicates that Russian forces seek to encircle Bakhmut by cutting off Ukrainian forces' access to the E40. Similarly, the report of a Russian attack on Chasiv Yar indicates that Russian forces have likely advanced closer to the T0504 Kostyantynivka–Chasiv Yar–Bakhmut highway southwest of Bakhmut.\xa0\xa0Russian sources claimed that Wagner Group fighters took control of Krasna Hora and are fighting northeast of Bakhmut.[30]\xa0Russian milbloggers also claimed that Wagner Group forces established fire control over a section of the T0504 highway between Stupochky and Ivanivske.[31]\n\n[32]\xa0Former Russian officer and prominent milblogger Igor Girkin claimed that Russian forces did not advance near Avdiivka and took heavy losses.\n\n[33]\xa0Another milblogger claimed that fighting is ongoing in western Marinka (on the southwestern outskirts of Donetsk City) and that unspecified elements of the Southern Military District (SMD) advanced through urban areas of Marinka on February 8.\n\n[34]\xa0The milblogger also stated that Russian forces were able to gain a foothold in positions near a tire repair plant in Marinka.\n\n[35]\xa0Videos posted by milbloggers on February 8 reportedly show SMD tank units attacking a Ukrainian position in Marinka and Russian tanks operating in western Marinka.\n\n[36]\xa0Former Deputy LNR Interior Minister Vitaly Kiselev posted a video on February 8 purportedly showing Russian elements of the 150th Motorized Rifle Division (8th Combined Arms Army, SMD) attacking Marinka and claimed that Russian forces had cleared all Ukrainian fortifications there.\n\n[37]\xa0The deployment of valuable Russian conventional military units (as opposed to DNR proxy forces) in the area is notable, if confirmed. Girkin, however, claimed that the situation in Marinka has not changed and continues at a sluggish pace.\n\n[38]\n\n[39]\xa0Russian sources made conflicting claims about the status of operations in this area. One milblogger claimed that fierce fighting is ongoing near Vuhledar (30km southwest of Donetsk City), while other milbloggers stated that there is no active fighting in the area.\n\n[40]\n\n[41]\n\n[42]\n\nSupporting Effort—Southern Axis\xa0(Russian objective: Maintain frontline positions and secure rear areas against Ukrainian strikes)\n\nRussian and Ukrainian forces reportedly continued small scale skirmishes and reconnaissance activity in the Dnipro River delta and on the Kinburn Spit on February 8. The United Kingdom Ministry of Defense (UK MoD) reported that Russian forces are using small boats to try to maintain a presence on islands in the Dnipro River delta south of Kherson City and that Ukrainian forces have deployed long-range artillery to strike several Russian outposts on the islands.[43]\xa0The UK MoD reported that Russian and Ukrainian forces have likely deployed small groups on the Kinburn Spit in Mykolaiv Oblast, aiming to control the Dnipro Gulf.[44]\xa0Ukraine's Southern Operational Command Spokesperson Natalia Humenyuk previously stated that Russian forces are increasing the number of reconnaissance and sabotage attempts in the area of the Dnipro River delta as part of an information operation to create a perceived threat against Kherson City.[45]\n\nRussian forces continue to construct defensive fortifications in Zaporizhia Oblast. Satellite imagery collected between January 26 and February 7 shows Russian forces expanding trench and field fortifications near Tarasivka, Zaporizhia Oblast.[46]\xa0Russian forces likely constructed these fortifications to further strengthen Russian positions along the T0401 highway between Polohy and Tokmak.\xa0\xa0Russian forces are likely estab

lishing long defensive lines along critical grounds lines of communicat ion (GLOCs) in Zaporizhia Oblast in preparation to defend against possi ble future Ukrainian counteroffensive operations along the Zaporizhia f rontline. However, ISW has not observed Russian forces constructing def enses intended to halt a cross-country Ukrainian attack on a large fron t, and defensive positions remain limited to main roads.\n\nRussian for ces continued routine fire west of Hulyaipole and in Dnipropetrovsk, Kh erson, and Mykolaiv Oblasts on February 8.[47]\xa0Ukrainian sources rep orted that Russian forces struck Kherson City and in the vicinity of Oc hakiv, Mykolaiv Oblast.[48]\n\nMobilization and Force Generation Effort s\xa0(Russian objective: Expand combat power without conducting general mobilization)\n\nRussian officials continued attempts to extend social benefits held by regular Russian servicemembers to volunteer formations serving in Ukraine. Russian Prime Minister Mikhail Mishustin stated on February 8 that the Russian government has prepared new measures to sup port volunteers, including increasing pensions and social assistance pa yments related to injuries and disabilities.[49]\xa0The Russian State D uma is reportedly drafting a bill to include provisions against discred iting volunteer detachments assisting the Russian military in Ukraine, as Wagner Group Financier Yevgeny Prigozhin previously demanded.[50]\xa 0The Kremlin is likely pursuing efforts to more formally recognize volu nteer formations in order to mitigate continued criticism of the Russia n Ministry of Defense (MoD) over the unclear status of volunteer format ions.[51]\n\nThe Wagner Group is reportedly resorting to more coercive tactics in its campaign to recruit prisoners, possibly in response to d eclining numbers of recruits since autumn 2022. Independent Russian out let\xa0Agentstvo\xa0reported on February 8 that Russian lawyers and hum an rights activists stated that Wagner Group representatives and Russia n Ministry of Internal Affairs and Federal Security Service (FSB) offic ials are threatening prisoners in Samara and Rostov oblasts, Krasnodar Krai, and the North Caucasus with new criminal cases if they refuse to volunteer with the Wagner Group in Ukraine.[52]\xa0One of the lawyers r eportedly stated that fewer convicts have agreed to volunteer with the Wagner Group in an unspecified recent period because of information abo ut its high casualties, supporting ISW’s previous assessment that Russi an convicts’ resistance may have caused a decline in the Wagner Group’s prison recruitment campaign.[53]\xa0The Wagner Group will likely contin ue these more coercive practices as it seeks to replenish its forces in Ukraine with more convict recruits following months of highly attrition al human wave attacks in eastern Ukraine.\n\nRussian officials continue to promote incremental efforts to fix longstanding personnel issues ass ociated with mobilization. Russian Deputy Chairman of the Federation Co uncil (and head of the mobilization working group) Andrey Turchak claim ed that the mobilization working group has received appeals from 22,000 Russian servicemembers and their family members since holding its first meeting on December 29, 2022, addressing issues like the incorrect accr ual of payments and the wrongful mobilization of fathers with many chil dren who should be exempt.[54]\xa0Turchak stated that the working group has heavily focused on solving poor recordkeeping issues through effort s to digitize military registration information from military recruitme nt offices.[55]\xa0Turchak claimed that the working group sent a report to Russian President Vladimir Putin with recommendations to establish c omprehensive rehabilitation centers, a minimum set of measures to suppo rt family members, a reduced term for recognizing a Russian soldier as missing, and a guarantee for receiving pensions.[56]\xa0These proposals and efforts are likely meant primarily to placate ultranationalist figu res that criticized the numerous issues associated with mobilization an d hedge against further domestic discontent ahead of a likely second wa ve of mobilization.\n\nActivity in Russian-occupied Areas\xa0(Russian o bjective: consolidate administrative control of and annexed areas; forc

ibly integrate Ukrainian civilians into Russian sociocultural, economi
c, military, and governance systems)\n\nRussian occupation authorities
are continuing efforts to increase connectivity between Russia and sout
hern Ukraine. Kherson Occupation Head Vladimir Saldo claimed on Februar
y 8 that Russian occupation authorities have approved design and resear
ch works on a new highway that will run from Crimea, north of the Sea o
f Azov, to Rostov-on-Don, Russia.[57]\xa0Saldo also claimed that the co
nstruction of a new town in the Arabat Spit in northeast Crimea has beg
un.[58]\xa0ISW has previously assessed that Russian occupation authorit
ies likely seek to increase the population in the deep rear of occupied
territories to strengthen production capabilities and support logistics
related to Russia's invasion of Ukraine.[59]\n\nRussian occupation auth
orities continue to lean on patronage-like partnerships with Russian fe
deral subjects to restore infrastructure in occupied territories. Donet
sk People's Republic (DNR) Head Denis Pushilin claimed on February 8 th
at he held a meeting with Sakhalin Oblast Governor Valery Limarenko in
which they discussed Sakhalin Oblast's plans to help repair kindergarte
ns, stadiums, schools, and playgrounds in occupied Shakhtarsk, Donetsk
Oblast.[60]\xa0Luhansk People's Republic (LNR) Head Leonid Pasechnik he
ld a meeting with Voronezh Oblast Governor Aleksandr Gusev on February
8 during which Gusev claimed that Voronezh Oblast hopes to develop occu
pied Luhansk Oblast to not only extract raw materials, but also to deve
lop a processing industry.[61]\xa0Gusev claimed that Voronezh Oblast wi
ll double the amount of aid it previously provided occupied Luhansk Obl
ast in 2022 to bring living standards in occupied Luhansk Oblast to tho
se of Russia's "national" level.[62]\n\nSignificant activity in Belarus
\xa0(ISW assesses that a Russian or Belarusian attack into northern Ukr
aine in early 2023 is extraordinarily unlikely and has thus restructure
d this section of the update. It will no longer include counter-indicat
ors for such an offensive.\n\nISW will continue to report daily observe
d Russian and Belarusian military activity in Belarus, but these are no
t indicators that Russian and Belarusian forces are preparing for an im
minent attack on Ukraine from Belarus. ISW will revise this text and it
s assessment if it observes any unambiguous indicators that Russia or B
elarus is preparing to attack northern Ukraine.)\n\nBelarusian airborne
forces may be conducting tactical force-on-force exercises with Russian
airborne elements in Belarus. The Belarusian Ministry of Defense announ
ced on February 8 that unspecified airborne infantry companies — likely
of the Belarusian 38th Air Assault Brigade — conducted a force-on-force
company tactical exercise at the Brest Training Ground, emphasizing usi
ng unmanned aerial vehicles, urban warfare, small unit tactics, and tac
tical medicine.[63]\xa0It is unclear if Russian airborne forces partici
pated in this exercise. The Belarusian 38th Air Assault Brigade has his
torically conducted joint exercises with elements of the Russian 76th A
ir Assault Division, 106th Airborne Division, and the 31st Air Assault
Brigade – all units that have taken casualties in Ukraine and require r
egeneration.[64]\n\nBelarusian maneuver elements continue conducting ex
ercises in Belarus. Unspecified elements of the Belarusian 19th Separat
e Guards Mechanized Brigade conducted tactical readiness exercises at t
he Lepelsky Training Ground in Vitebsk Oblast, Belarus, on February 8.
[65]\n\nNote: ISW does not receive any classified material from any sou
rce, uses only publicly available information, and draws extensively on
Russian, Ukrainian, and Western reporting and social media as well as c
ommercially available satellite imagery and other geospatial data as th
e basis for these reports. References to all sources used are provided
in the endnotes of each update.\n\n[1]\xa0https://isw.pub/UkrWar02062
3;\xa0https://isw.pub/UkrWar020423; https://isw.pub/UkrWar020223\n\n[2]
\xa0https://t.me/DeepStateUA/15451\n\n[3]\xa0https://www.understandingw
ar.org/backgrounder/russian-offensive-campaign...\n\n[4]\xa0https://t.m
e/rybar/43387;\xa0https://t.me/notes_veterans/7845;\xa0https://t....\n

\n[5]https://www.understandingwar.org/sites/default/files/Russian%20Operation...\n\n[6]\xa0https://www.understandingwar.org/backgrounder/russian-offensive-campaign...\n\n[7]\xa0https://t.me/ostorozhno_novosti/14167\n\n[8]\xa0https://isw.pub/UkrWar020323\xa0;\xa0https://isw.pub/UkrWar020423\n\n[9]\xa0https://t.me/ostorozhno_novosti/14167\n\n[10]\xa0https://podolyaka\xa0dot ru/2023/02/08/zayavleniya-premer-ministra-rf-mihaila-mishustina-o-podderzhke-uchastnikov-svo-i-voennoy-promyshlennosti/;\xa0https://stolica-s\xa0dot su/archives/366231;\xa0https://t.me/rybar/43402\n\n[11]\xa0https://podolyaka\xa0dot ru/2023/02/08/zayavleniya-premer-ministra-rf-mihaila-mishustina-o-podderzhke-uchastnikov-svo-i-voennoy-promyshlennosti/;\xa0https://stolica-s\xa0dot su/archives/366231;\xa0https://t.me/rybar/43402\n\n[12]\xa0https://isw.pub/UkrWar011823\xa0;\xa0https://isw.pub/UkrWar010723\xa0;\n\n[13]\xa0https://isw.pub/UkrWar020623\xa0;\xa0https://isw.pub/UkrWar020123\n\n[14]\xa0https://www.gov.uk/government/news/new-sanctions-target-putins-war-machi...\n\n[15]\xa0https://www.cbsnews.com/news/ukraine-war-russia-iranian-drones-us-made-t...\n\n[16]\xa0https://suspilne\xa0dot media/amp/378863-de-okupanti-posiluut-prisutnist-na-harkivsini-dani-sinegubova/\n\n[17]\xa0https://t.me/kommunist/15598; https://understandingwar.org/backgrounder/russian-offensive-campaign-ass...\n\n[18]https://www.facebook.com/GeneralStaff.ua/posts/pfbid0FuH223o7wLNSiJSNdCX...\n\n[19]\xa0https://t.me/strelkovii/3896\n\n[20]\xa0https://t.me/luhanskaVTSA/8438\n\n[21]https://www.facebook.com/GeneralStaff.ua/posts/pfbid0FuH223o7wLNSiJSNdCX...\n\n[22]\xa0https://t.me/russkiy_opolchenec/35783; https://t.me/RVvoenkor/37711\n\n[23]\xa0https://t.me/rybar/43387;\xa0https://t.me/notes_veterans/7845\n\n[24]\xa0https://t.me/wargonzo/10782\n\n[25]https://www.facebook.com/GeneralStaff.ua/posts/pfbid0FuH223o7wLNSiJSNdCX...\n\n[26]\xa0https://t.me/RKadyrov_95/3332\n\n[27]\xa0https://twitter.com/fdov21/status/1623368452667805701https://twitter.com/SerDer_Daniels/status/1623295739890630657;\xa0https://...\n\n[28]\xa0https://twitter.com/Militarylandnet/status/1623071883988987905; https://twitter.com/EerikMatero/status/1623076900548517892; https://twitter.com/neonhandrail/status/1623206937134497792; https://twitter.com/neonhandrail/status/1623207358888558593\n\n[29]https://www.facebook.com/GeneralStaff.ua/posts/pfbid0FuH223o7wLNSiJSNdCX...\n\n[30]\xa0https://t.me/wargonzo/10773;\xa0https://t.me/strelkovii/3896\n\n[31]\xa0https://t.me/DonbassYasinovatayanaliniiOgnia/36445;\xa0https://t.me/Neofic...\n\n[32]https://www.facebook.com/GeneralStaff.ua/posts/pfbid0FuH223o7wLNSiJSNdCX...\n\n[33]\xa0https://t.me/strelkovii/3896\n\n[34]\xa0https://t.me/rybar/43405\n\n[35]\xa0https://t.me/rybar/43405\n\n[36]\xa0https://t.me/boris_rozhin/77568;\xa0https://t.me/sashakots/38439\n\n[37]\xa0https://t.me/kommunist/15635\n\n[38]\xa0https://t.me/strelkovii/3896\n\n[39]https://www.facebook.com/GeneralStaff.ua/posts/pfbid0FuH223o7wLNSiJSNdCX...\n\n[40]\xa0https://t.me/boris_rozhin/77574;\xa0https://t.me/wargonzo/10773\n\n[41]\xa0https://t.me/strelkovii/3896\n\n[42]\xa0https://t.me/Bratchuk_Sergey/29230\n\n[43]\xa0https://twitter.com/DefenceHQ/status/1623199796352745475/photo/1\n\n[44]\xa0https://twitter.com/DefenceHQ/status/1623199796352745475/photo/1\n\n[45]\xa0https://armyinform dot\xa0com.ua/2023/02/01/zbilshennya-kilkosti-rozviduvalnyh-grup-voroga-v-gyrli-dnipra-mozhe-buty-oznakoyu-nagnitannya-sytuacziyi-gumenyuk/\n\n[46]\xa0https://twitter.com/bradyafr/status/1623082928283746304?s=20&t=ETx-WeYab...\n\n[47]\n\nhttps://t.me/mykolaivskaODA/4236\xa0;\n\nhttps://www.facebook.com/GeneralStaff.ua/posts/pfbid0FuH223o7wLNSiJSNdCX...\xa0;\n\nhttps://www.facebook.com/GeneralStaff.ua/posts/pfbid02kL8XZwXNsSUphpcF5S...\xa0;\n\nhttps://t.me/khersonskaODA/3607\xa0;\n\nhttps://t.me/khersonskaODA/3616;\n\nhttps://t.me/khersonskaODA/3613;\n\nhttps://t.me/khersonskaODA/3615\xa0;\n\nhttps://www.facebook.com/sergey.khlan/posts/pfbid02L4QqnKMLM3QLzY1pvQrr5...\xa0;\n\nhttps://t.me/mykolaivskaODA/4236\xa0\xa0;\n\nhttps://t.me/zoda_gov_ua/16505\xa0\xa0;\n\nhttps://t.me/vilkul/2680\xa0\xa0;\n\nhttps://t.me/Yevtushenko_E/2419\n\n[48]\xa0https://t.me/mykolaivskaODA/4236\xa0;\x

a0\xa0https://www.facebook.com/GeneralStaff.ua/posts/pfbid0FuH223o7wLNS
iJSNdCX...\xa0;\xa0\xa0https://www.facebook.com/GeneralStaff.ua/posts/p
fbid02kL8XZwXNsSUphpcF5S...\xa0;\xa0https://t.me/khersonskaODA/3607\n\n
[49]\xa0https://podolyaka dot ru/2023/02/08/zayavleniya-premer-ministra
-rf-mihaila-mishustina-o-podderzhke-uchastnikov-svo-i-voennoy-promyshle
nnosti/; https://stolica-s dot su/archives/366231; https://t.me/rybar/4
3402\n\n[50]\xa0https://isw.pub/UkrWar020723; https://t.me/concordgroup
_official/425; https://t.me/Prigozhin_hat/2612\n\n[51]\xa0https://www.u
nderstandingwar.org/backgrounder/russian-offensive-campaign...\n\n[52]
\xa0https://t.me/agentstvonews/2482\n\n[53]\xa0https://t.me/agentstvone
ws/2482\xa0;\xa0https://isw.pub/UkrWar020123\n\n[54]\xa0https://t.me/tu
rchak_andrey/1032\xa0;\xa0\xa0https://isw.pub/UkrWar122922\n\n[55]\xa0h
ttps://t.me/turchak_andrey/1032\n\n[56]\xa0https://t.me/turchak_andrey/
1032\n\n[57]\xa0https://t.me/SALDO_VGA/363;\xa0https://t.me/SALDO_VGA/3
65\n\n[58]\xa0https://t.me/SALDO_VGA/363;\xa0https://t.me/SALDO_VGA/365
\n\n[59]\xa0https://www.understandingwar.org/backgrounder/russian-offen
sive-campaign...\n\n[60]\xa0https://t.me/pushilindenis/3165\n\n[61]\xa0
https://t.me/glava_lnr_info/745;\xa0https://t.me/glava_lnr_info/748\n\n
[62]\xa0https://t.me/glava_lnr_info/745;\xa0https://t.me/glava_lnr_inf
o/748\n\n[63]\xa0https://t.me/modmilby/22455; https://t.me/modmilby/224
19\n\n[64]\xa0https://www.osw.waw.pl/sites/default/files/OSW-Report_Rus
sia%E2%80%99s-B...\n\n[65]\xa0https://t.me/modmilby/22470\n\nTags\n\nUk
raine Project\n\nFile Attachments:\n\nZaporizhia Battle Map Draft Febru
ary 08,2023.png\n\nKherson-Mykolaiv Battle Map Draft February 08,2023.p
ng\n\nDonetsk Battle Map Draft February 08,2023.png\n\nDraftUkraineCoTF
ebruary08,2023.png\n\nKharkiv Battle Map Draft February 08,2023.png\n\n
1400 16th Street NW, Suite 515 Washington, DC 20036', metadata={'sourc
e': 'https://www.understandingwar.org/backgrounder/russian-offensive-ca
mpaign-assessment-february-8-2023'}),
 Document(page_content='Skip to main content\n\nSearch form\n\nHome\n\nW
ho We Are\n\nResearch\n\nPublications\n\nGet Involved\n\nPlanned Giving
\n\nDonate\n\nRussian Offensive Campaign Assessment, February 9, 2023\n
\nFeb 9, 2023 - Press ISW\n\nDownload the PDF\n\nKarolina Hird, Riley Ba
iley, George Barros, Nicole Wolkov, and Frederick W. Kagan\n\nFebruary
9, 7:30 pm ET\n\nClick here to see ISW’s interactive map of the Russian
invasion of Ukraine. This map is updated daily alongside the static maps
present in this report.\n\nWagner Group financier Yevgeny Prigozhin anno
unced on February 9 that the Wagner Group has entirely stopped recruitin
g prisoners.\xa0In a response to a press comment, Prigozhin claimed that
Wagner’s recruitment of prisoners has "completely stopped" and that "all
obligations are being fulfilled" for those currently under Wagner’s empl
oy.[1]\xa0Prigozhin also absurdly claimed that over 10 million Americans
have applied to join Wagner.[2]\xa0The Wagner Group will likely continue
to recruit from prisons, albeit in a much more limited capacity. As ISW
has previously noted, Wagner’s recruitment of prisoners has slowed over
the last few months, an assessment confirmed by statistics by the Federa
l Penitentiary Service that show that decreases in the Russian prison po
pulation stabilized between November 2022 and January 2023.[3]\xa0This p
henomenon is consistent with the overall trend of conventional Russian t
roops slowly replacing the Wagner Group around Bakhmut, indicating that
Russian military command may be shifting away from its reliance on Wagne
r and therefore on using prisoners as cannon fodder.\n\nThe Kremlin cont
inues to pursue measures to gradually prepare Russia’s defense-industria
l base for a protracted war in Ukraine while avoiding wider economic mob
ilization.\xa0Russian President Vladimir Putin held a meeting with the S
upervisory Board of the Agency for Strategic Initiatives on February 9 a
nd instructed the agency to support federal subjects in developing the p
roduction of unmanned aircraft systems.[4]\xa0Deputy Chairman of the Rus
sian Security Council Dmitry Medvedev previously stated that Russian mil
itary manufacturers intend to expand the supply of reconnaissance and at

tack drones to support operations in Ukraine, and Russian and Iranian of ficials are reportedly planning to build a factory in Russia to manufact ure 6,000 drones "in the coming years."[5]\n\nMedvedev visited a tank ma nufacturing plant in Omsk Oblast on February 9 and stated that Russia ne eds to increase the production of various armaments, including modern ta nks, in response to Western military assistance to Ukraine.[6]\xa0Dutch open-source group Oryx reported that Russian forces have lost 1,012 dest royed tanks in Ukraine with an additional 546 tanks captured by Ukrainia n forces.[7]\xa0Oryx reported that these combined losses represent rough ly half the tanks that Russian forces committed to Ukraine at the start of the invasion.[8]\xa0Fifteen hundred tanks are enough to equip more th an 15 tank regiments or brigades or about 150 battalion tactical groups. [9]\xa0The Russian military needs to quickly replenish these tank losses to maintain the ability to conduct large-scale mechanized maneuver warfa re ahead of a likely increased pace of offensive operations in eastern U kraine. Medvedev likely framed his calls for increased production as a r esponse to Western military assistance to obscure the fact that substant ial military equipment losses are driving the need for increased product ion. The Kremlin's efforts to gradually prepare Russia's defense industr ial base for a protracted war while avoiding a wider mobilization of the Russian economy continue to be incompatible with the scale of the war th at the Russian military is fighting in Ukraine and the scale of Russian military equipment losses.\n\nA prominent Wagner-linked Russian milblogg er called for the dismissal of Russian Defense Minister Sergei Shoigu ov er a Russian military uniform procurement scandal.\xa0Many prominent Rus sian military bloggers harshly criticized Shoigu and the Russian Ministr y of Defense (MoD) over news that the 22-year-old son of the Russian Dep uty Head of the Federal Agency for State Property Management won a cont ract to supply the Russian military with new uniforms.[10]\xa0The milbl oggers argued that the new uniforms are of inferior quality and overpri ced (costing about 130,000-210,000 rubles or $1,780 - $2,875 per unifor m) and are part of a petty corruption scheme to enrich the families of Russian defense officials. The Grey Zone Telegram channel—a prominent W agner Group-affiliated milblogger — wrote an explicative-laden rant to its 426,000 subscribers that Shoigu has lost credibility in front of th e Russian nation and that Russian President Vladimir Putin can amend th e situation by firing Shoigu, Shoigu's "entourage" in the Russian Gener al Staff and banning Shoigu and his associates from all Russian militar y affairs.[11]\xa0This is the latest episode in a string of events that has prompted Russian military blogger communities to attack the Russian MoD and senior Kremlin officials for petty corruption and ineptitude re sulting in battlefield failures and worse quality of life for average R ussian soldiers.[12]\n\nThe Kremlin continues to show that it is unwill ing to curb divisive rhetoric from ultranationalist pro-war figures.\xa 0Chechen Republic head Ramzan Kadyrov publicly sparred with Duma Deputy General Viktor Sobolev following Sobolev's criticism of Kadyrov's state ments on grooming standards in the Russian military being discriminator y against Muslims and calls for the Russian military to fight satanism in Poland.[13]\xa0Kremlin spokesperson Dmitry Peskov stated on February 9 that the Kremlin is "not participating in this controversy and would not like to give any assessments" about it.[14]\xa0The Kremlin will con tinue to tolerate divisive rhetoric from ultranationalist figures as it seeks to appeal to the wider pro-war community.\n\nKey Takeaways\n\nWag ner Group financier Yevgeny Prigozhin announced that the Wagner Group h as entirely stopped recruiting prisoners.\n\nThe Kremlin continues to p ursue measures to gradually prepare Russia's defense industrial base fo r a protracted war in Ukraine.\n\nA prominent Wagner-linked Russian mil blogger called for the dismissal of Russian Defense Minister Sergey Sho igu over a Russian military uniform procurement scandal.\n\nThe Kremlin continues to illustrate that it is unwilling to curb divisive rhetoric

from ultranationalist pro-war figures.\n\nRussian forces continued offe
nsive operations along the Svatove-Kreminna line.\n\nRussian forces con
ducted limited ground attacks in western Donetsk Oblast and the Avdiivk
a-Donetsk City area and continued offensive operations around Bakhmu
t.\n\nRussian forces conducted a limited ground attack in Zaporizhia Ob
last.\n\nRussian sources claimed that the Russian military integrated a
Donetsk People's Republic (DNR) volunteer formation into the Russian Ar
med Forces.\n\nRussian sources claimed that Russian authorities detaine
d a Ukrainian sabotage and reconnaissance group attempting to assassina
te Russian occupation officials.\n\nWe do not report in detail on Russi
an war crimes because those activities are well-covered in Western medi
a and do not directly affect the military operations we are assessing a
nd forecasting. We will continue to evaluate and report on the effects
of these criminal activities on the Ukrainian military and population a
nd specifically on combat in Ukrainian urban areas. We utterly condemn
these Russian violations of the laws of armed conflict, Geneva Conventi
ons, and humanity even though we do not describe them in these report
s.\n\nRussian Main Effort—Eastern Ukraine (comprised of two subordinate
main efforts)\n\nRussian Subordinate Main Effort #1—Capture the remaind
er of Luhansk Oblast and push westward into eastern Kharkiv Oblast and
encircle northern Donetsk Oblast\n\nRussian Subordinate Main Effort #2—
Capture the entirety of Donetsk Oblast\n\nRussian Supporting Effort—Sou
thern Axis\n\nRussian Mobilization and Force Generation Efforts\n\nActi
vities in Russian-occupied Areas\n\nRussian Main Effort—Eastern Ukraine
\n\nRussian Subordinate Main Effort #1- Luhansk Oblast (Russian objecti
ve: Capture the remainder of Luhansk Oblast and continue offensive oper
ations into eastern Kharkiv Oblast and northern Donetsk Oblast)\n\nISW
continues to assess that Russia\'s most likely course of action (MLCOA)
is an imminent offensive effort in Luhansk Oblast and is therefore adju
sting the structure of the daily campaign assessments. We will no longe
r include the Eastern Kharkiv and Western Luhansk Oblast area as part o
f Ukrainian counteroffensives and will assess this area as a subordinat
e part of the Russian main effort in Eastern Ukraine. The assessment of
Luhansk Oblast as part of the Russian main effort does not preclude the
possibility of continued Ukrainian counteroffensive actions here or any
where else in theater in the future. ISW will report on Ukrainian count
eroffensive efforts as they occur.\n\nRussian forces continued offensiv
e actions along the Svatove-Kreminna line on February 9. The Ukrainian
General Staff reported that Ukrainian troops repelled a Russian attack
near Stelmakhivka, 15km west of Svatove.[15]\xa0Russian milbloggers cir
culated footage reportedly of elements of the 3rd Motor Rifle Division
(20th Combined Arms Army, Western Military District) correcting artille
ry strikes in an unidentified sector of the Svatove-Kreminna line.[16]
\xa0The Russian Ministry of Defense (MoD) claimed that \xa0Central Mili
tary District elements are operating in the Lyman direction (the area w
est of Kreminna) and using TOS-1 thermobaric multiple rocket launch sys
tems.[17]\xa0The commitment of a military district-level asset such as
the TOS-1 to the Kreminna area suggests that the Russian MoD is priorit
izing this axis. Widely circulated social media footage posted on Febru
ary 9 additionally shows a Ukrainian strike on a Russian BMPT Terminato
r armored fighting vehicle about 8km south of Kreminna, indicating that
the Russian command is committing new equipment to this area of the fro
nt.[18]\xa0The Ukrainian General Staff noted that Ukrainian troops repe
lled a Russian attack near Bilohorivka.[19]\xa0A Russian milblogger cla
imed that Russian troops are conducting offensive operations north of B
ilohorivka and attacked along the Shepilove-Dibrova line, about 5km sou
th of Kreminna.[20]\n\nRussian Subordinate Main Effort #2—Donetsk Oblas
t\xa0(Russian objective: Capture the entirety of Donetsk Oblast, the cl
aimed territory of Russia's proxies in Donbas)\n\nRussian forces contin
ued ground attacks around Bakhmut on February 9. The Ukrainian General

Staff reported that Ukrainian troops repelled Russian attacks on Bakhmu t itself; northeast of Bakhmut near Vyimka (22km northeast) and Fedoriv ka (15km northeast); north of Bakhmut near Krasna Hora (4km north) and Paraskoviivka (5km north); and west of Bakhmut near Ivanviske (5km wes t) and Chasiv Yar (10km west).[21]\xa0A Russian milblogger remarked tha t Russian troops have recently changed their tactics in the Bakhmut are a and are focusing less on frontal assaults on small settlements and mo re on interdicting Ukrainian ground lines of communication (GLOCs) into Bakhmut along the E40 Bakhmut–Slovyansk and T0504 Kostyantynivka–Chasiv Yar–Bakhmut highways.[22]\xa0This observation is consistent with the Uk rainian General Staff report of Russian attacks towards Chasiv Yar and Ivanivske, both critical settlements along the T0504. Other Russian mil bloggers similarly claimed that Wagner Group forces are pushing towards Ivanivske and attacking along the E40 near Orikhovo-Vasylivka (10km nor thwest of Bakhmut) and Dubovo-Vasylivka (5km northwest of Bakhmut).[23] \xa0Russian sources claimed that Wagner Group forces are additionally a ttacking toward Krasna Hora from three sides and that Ukrainian troops are close to withdrawing from the settlement.[24]\n\nRussian forces con ducted limited ground attacks in the Avdiivka–Donetsk City area on Febr uary 9. The Ukrainian General Staff reported that Ukrainian forces repe lled Russian assaults near Avdiivka, north of Avdiivka near Novokalynov e, and along the western outskirts of Donetsk City near Vodyane, Pervom aiske, Vesele, Krasnohorivka, and Marinka.[25]\xa0A Russian milblogger claimed that Russian forces on the northwestern outskirts of Donetsk Ci ty resumed offensives near Krasnohorivka, advanced near Vodyane and Opy tne, failed to break through near Pervomaiske, and continued attacks in western Marinka.[26]\xa0Social media footage published on February 9 pu rportedly shows elements of the 5th Brigade of the 1st Army Corps (forc es of the Donetsk People’s Republic) attacking Ukrainian positions near Marinka.[27]\n\nRussian forces conducted limited ground attacks in west ern Donetsk Oblast on February 9. The Ukrainian General Staff reported that Ukrainian forces repelled Russian assaults near Bohoyavlenka (25km southwest of Donetsk City) and Prechystivka (38km southwest of Donetsk City).[28]\xa0A Russian milblogger claimed that Russian forces resumed assault operations on the outskirts of Vuhledar.[29]\xa0A Ukrainian res erve officer also reported that the majority ethnic Tatar volunteer bat talion "Alga" of the 72nd Motorized Rifle Brigade (3rdArmy Corps) fough t near Vuhledar on February 6.[30]\xa0The reserve officer suggested tha t the use of volunteer battalions in this area indicates that the 155th and 40th Naval Infantry Brigades, which were previously active in the a rea, sustained insurmountable losses and are being replaced by other fo rmations.[31]\xa0Recently posted footage from the Vuhledar area shows a defeated Russian mechanized formation of the 155th Naval Infantry Briga de that lost 13 main battle tanks and 12 BMP infantry fighting vehicles in a single engagement – about half a Russian tank battalion.[32]\xa0Th e footage shows the Russian formation driving in a column displaying po or tactics and a lack of learning from previous Russian tactical failur es.[33]\xa0Separate drone footage published on February 8 shows Ukraini an forces striking Russian forces approaching Vuhledar.[34]\xa0Geolocat ed footage published February 7 also shows reported elements of the 36t h\xa0Separate Guards Motorized Rifle Brigade (29th Combined Arms Army, Eastern Military District) striking Ukrainian positions on the eastern outskirts of Vuhledar.[35]\n\nSupporting Effort—Southern Axis\xa0(Russi an objective: Maintain frontline positions and secure rear areas agains t Ukrainian strikes)\n\nRussian forces conducted a limited ground attac k in Zaporizhia Oblast on February 9. The Ukrainian General Staff repor ted that Ukrainian forces repelled a Russian assault near Novoandriivk a, Zaporizhia Oblast.[36]\xa0Russian forces additionally continued rout ine fire west of Hulyaipole and in Dnipropetrovsk and Kherson oblasts o n February 9.[37]\xa0Ukrainian sources reported that Russian forces str

uck Kherson City and Nikopol, Dnipropetrovsk Oblast.[38]\n\nMobilizatio
n and Force Generation Efforts\xa0(Russian objective: Expand combat pow
er without conducting general mobilization)\n\nRussian sources claimed
that the Russian military integrated a Donetsk People's Republic (DNR)
volunteer formation into the Russian Armed Forces. A Russian milblogger
claimed that the former DNR Vostok volunteer battalion is now formally
the 11th Regiment of the Russian Armed Forces.[39]\xa0The DNR Vostok vo
lunteer battalion commander, a notable Russian milblogger, has not addr
essed the reported integration. The Russian milblogger did not specify
what higher formation the 11th regiment is subordinated to, but it is p
ossible that it is subordinated to the Southern Military District which
formally controls the DNR 1st Army Corps. This regiment may be assigned
to a new unspecified Russian maneuver division. ISW previously assessed
that the Russian Ministry of Defense (MoD) appears to be rushing to int
egrate irregular conventional forces into more traditional structures a
nd will likely struggle to correct these formations' poor effectiveness
during integration efforts.[40]\n\nA Russian military court reportedly
ruled that Russian commanders can legally refuse to release servicememb
ers from service at the end of their contracts. A Russian media outlet
reported on February 8 that a Perm Oblast military garrison court ruled
in favor of Russian commanders in a lawsuit filed by a serviceman who c
laimed that the commanders refused to release him from service when his
contract ended in September 2022.[41]\xa0The court reportedly argued th
at the partial mobilization decree established an exhaustive list of gr
ounds for release that did not include the expiration of a contract, an
d therefore, concluded that there were no grounds for recognizing the a
ctions of the commanders as illegal.[42]\xa0ISW continues to assess tha
t the Kremlin will not formally rescind the partial mobilization decree
to legally justify the continued service of mobilized personnel indefin
itely.\n\nA volunteer battalion affiliated with a Russian occupation of
ficial reportedly deployed to frontline positions in Ukraine. Zaporizhi
a Oblast occupation deputy Vladimir Rogov claimed on February 9 that el
ements of Zaporizhia Oblast occupation head Yevgeny Balitsky's Sudaplat
ov volunteer battalion have deployed to frontline positions in an unspe
cified area of Ukraine.[43]\xa0Russian sources have previously claimed
that Turkish, Swedish, and Serbian volunteers are serving in the volunt
eer battalion.[44]\xa0ISW continues to assess that this volunteer batta
lion will likely face significant command and control challenges if the
se claims are true.\n\nActivity in Russian-occupied Areas\xa0(Russian o
bjective: consolidate administrative control of and annexed areas; forc
ibly integrate Ukrainian civilians into Russian sociocultural, economi
c, military, and governance systems)\n\nThe Kremlin continues to priori
tize the development of occupied territories in Ukraine. Russian Presid
ent Vladimir Putin met with the Supervisory Board of the Agency for Str
ategic Initiatives on February 9 and instructed the agency to focus on
the development and implementation of socio-economic development plans
for occupied Luhansk, Donetsk, Zaporizhia, and Kherson oblasts.[45]\xa0
Putin stated that Russia needs to develop the occupied territories with
in the current decade to foster social commitment to Russia.[46]\xa0Don
etsk People's Republic head Denis Pushilin met with Russian Energy Mini
ster Nikolai Shulginov on February 9 to discuss a new fuel and energy c
omplex in Donetsk Oblast.[47]\xa0The Kremlin and Russian occupation off
icials likely believe that rapid economic development will promote wide
spread pro-Russian sentiments in occupied territories.\n\nRussian offic
ials continue to pursue the deportation of residents and children from
occupied territories through various schemes. Zaporizhia Oblast occupat
ion head Yevgeny Balitsky stated on February 9 that Russian medical ser
vices are taking women and newborns with diseases that cannot be treate
d in Zaporizhia Oblast to Russia for perinatal care.[48]\xa0Balitsky al
so stated that Russian doctors and the Russian Ministry of Health are p

roviding obstetric services in occupied Zaporizhia Oblast.[49]\xa0Both these obstetric measures are likely meant to support ongoing Russian efforts to deport children and residents from occupied territories under the guise of medical relocation schemes.[50]\xa0Russian Agency for Strategic Initiatives member Svetlana Chupsheva stated in a meeting with Putin that the agency helped relocate dozens of residents from occupied territories to Russia under the medical relocation scheme.[51]\xa0Chupsheva also stated that the Agency for Strategic Initiatives plans to accept 5,000 children from Donetsk Oblast into a programming course that may set conditions for the relocation of these children to Russia.[52]\xa0ISW continues to assess that these schemes are likely part of a wider ethnic cleansing effort.\n\nRussian sources claimed on February 9 that Russian authorities detained members of a Ukrainian sabotage and reconnaissance group attempting to assassinate the Berdyansk deputy occupation head, the Berdyansk occupation traffic police deputy head, and the Berdyansk occupation commandant.[53]\xa0Russian sources also claimed that the Ukrainian sabotage and reconnaissance group was preparing an attack on a "We are Together with Russia" Center near Berdyansk.[54]\xa0ISW has previously assessed that the Kremlin likely founded, coordinates, and promotes the "We Are Together with Russia" organization to create a facade of public support for the annexation and integration of occupied Ukrainian oblasts into Russia.[55]\n\nSignificant activity in Belarus\xa0 (ISW assesses that a Russian or Belarusian attack into northern Ukraine in early 2023 is extraordinarily unlikely and has thus restructured this section of the update. It will no longer include counter-indicators for such an offensive.\n\nISW will continue to report daily observed Russian and Belarusian military activity in Belarus, but these are not indicators that Russian and Belarusian forces are preparing for an imminent attack on Ukraine from Belarus. ISW will revise this text and its assessment if it observes any unambiguous indicators that Russia or Belarus is preparing to attack northern Ukraine.)\n\nBelarusian maneuver elements continue conducting exercises in Belarus. Unspecified elements of the Belarusian 6th\xa0Separate Guards Mechanized brigade conducted tactical live-fire exercises at the Gozhsky Training Ground in Grodno, Belarus, on February 9.[56]\xa0A tank battalion of the Belarusian 11th Separate Mechanized Brigade conducted live fire training with T-72 tanks at the Obuz-Lesnovsky Training Ground in Brest, Belarus, on February 9.[57]\n\nBelarus reportedly began a month-long training for reserve recruits on February 9. The Grey Zone Telegram channel reported on February 9 that Belarus' reserve recruits began a one-month-long training period.[58]\n\nNote: ISW does not receive any classified material from any source, uses only publicly available information, and draws extensively on Russian, Ukrainian, and Western reporting and social media as well as commercially available satellite imagery and other geospatial data as the basis for these reports. References to all sources used are provided in the endnotes of each update.\n\n[1]\xa0https://t.me/concordgroup_official/426\n\n[2]\xa0https://t.me/concordgroup_official/427\n\n[3]\xa0https://www.understandingwar.org/backgrounder/russian-offensive-campaign...\n\n[4]\xa0http://kremlin\xa0dot ru/events/president/news/70482\n\n[5]\xa0https://www.wsj.com/articles/moscow-tehran-advance-plans-for-iranian-des... ; https://tass dot\xa0ru/armiya-i-opk/16872633 ;\n\n[6]\xa0https://t.me/medvedev_telegram/265\xa0;\xa0https://www.reuters.com/world/eur...\n\n[7]\xa0https://www.cnn.com/2023/02/09/europe/1000-russian-tanks-destroyed-ukrai...\n\n[8]\xa0https://www.cnn.com/2023/02/09/europe/1000-russian-tanks-destroyed-ukrai...\n\n[9]\xa0https://www.businessinsider.com/captured-documents-say-elite-russian-unit-lost-tanks-kharkiv-ukraine-2022-5; https://rusi.org/explore-our-research/publications/commentary/getting-know-russian-battalion-tactical-group\n\nhttps://www.rbc dot ru/society/09/02/2023/63e3b0f79a7947b95a3c8cfc;\n\nhttps://t.me/rybar/43427;\n\nhttps://t.me/milinfolive/96749;\n\nhttps://t.me/rustroyk

a1945/8471;\n\nhttps://t.me/rustroyka1945/8475;\n\nhttps://t.me/rustroyka1945/8476;\n\nhttps://t.me/rustroyka1945/8477;\n\nhttps://t.me/rybar/43435;\n\nhttps://t.me/m0sc0wcalling/19454;\xa0https://t.me/m0sc0wcalling/19453;\n\nhttps://t.me/grey_zone/17119;\n\nhttps://t.me/grey_zone/17117;\n\nhttps://t.me/grey_zone/17116;\xa0https://t.me/grey_zone/17112;\n\nhttps://t.me/boris_rozhin/77624\n\n[11]\xa0https://t.me/grey_zone/17119; https://t.me/grey_zone/17117; https://t.me/grey_zone/17116; https://t.me/grey_zone/17112\n\n[12]\xa0https://www.understandingwar.org/backgrounder/russian-offensive-campaign... https://www.understandingwar.org/backgrounder/russian-offensive-campaign...\n\n[13]\xa0https://t.me/RKadyrov_95/3334;\xa0https://eng.kavkaz-uzel\xa0dot eu/articles/62013/;\xa0https://eng.kavkaz-uzel\xa0dot eu/articles/62012/\n\n[14]\xa0https://ria\xa0dot ru/20230209/kadyrov-1850804605.html\n\n[15]\xa0https://www.facebook.com/GeneralStaff.ua/posts/pfbid02LGCyzg9CHrzXTN98Pj...\n\n[16]\xa0https://t.me/rybar/43443; https://t.me/dva_majors/8915\n\n[17]\xa0https://t.me/mod_russia/24064\n\n[18]\xa0https://m.facebook.com/story.php?story_fbid=pfbid0PHvkPzQjU7XJexyyQKsrJ4...\n\n[19]\xa0https://www.facebook.com/GeneralStaff.ua/posts/pfbid02LGCyzg9CHrzXTN98Pj...\n\n[20]\xa0https://t.me/wargonzo/10787\n\n[21]\xa0https://www.facebook.com/GeneralStaff.ua/posts/pfbid02LGCyzg9CHrzXTN98Pj...\n\n[22]\xa0https://t.me/milchronicles/1542\n\n[23]\xa0https://t.me/wargonzo/10787;\xa0https://t.me/readovkanews/52337\n\n[24]\xa0https://t.me/milinfolive/96786\xa0; https://t.me/rlz_the_kraken/56339\n\n[25]\xa0https://www.facebook.com/GeneralStaff.ua/posts/pfbid02LGCyzg9CHrzXTN98Pj...\n\n[26]\xa0https://t.me/wargonzo/10787\n\n[27]\xa0https://t.me/nm_dnr/9867;\xa0https://t.me/boris_rozhin/77599\n\n[28]\xa0https://www.facebook.com/GeneralStaff.ua/posts/pfbid02LGCyzg9CHrzXTN98Pj...\n\n[29]\xa0https://t.me/wargonzo/10787\n\n[30]\xa0https://twitter.com/Tatarigami_UA/status/1623436287171776514?s=20&t=mlsf...\n\n[31]\xa0https://twitter.com/Tatarigami_UA/status/1623436287171776514?s=20&t=mlsf...\n\nhttps://twitter.com/UAWeapons/status/1623649601717772288;\n\nhttps://twitter.com/Osinttechnical/status/1623532179220500480;\n\nhttps://t.me/m0sc0wcalling/19484;\n\nhttps://twitter.com/fdov21/status/1623453631629467650?s=20&t=u8lPsoZX5qq...\n\nhttps://twitter.com/markito0171/status/1623418942025785344?s=20&t=_NUJZh...\n\nhttps://twitter.com/markito0171/status/1623426654004469760?s=20&t=_NUJZh...\n\nhttps://twitter.com/markito0171/status/1623598095924711425?s=20&t=_NUJZh...\n\nhttps://t.me/m0sc0wcalling/19484\n\n[33]\xa0https://twitter.com/UAWeapons/status/1623649601717772288;\xa0https://twitt...\n\n[34]\xa0https://twitter.com/Militarylandnet/status/1623386771282075676?s=20&t=dd...\n\n[35]\xa0https://t.me/rusichtank/94 ; https://twitter.com/3_bm15/status/1623670248699310080?s=20&t=APP4gCFnjvU... ; https://twitter.com/3_bm15/status/1623670378458411009?s=20&t=APP4gCFnjvU... ; https://twitter.com/GeoConfirmed/status/1623730058211778563?s=20&t=APP4g...\n\n[36]\xa0https://www.facebook.com/GeneralStaff.ua/posts/pfbid02LGCyzg9CHrzXTN98Pj...\n\n[37]\xa0https://www.facebook.com/GeneralStaff.ua/posts/pfbid02opH8aKBK88janDxszu... https://twitter.com/IntelCrab/status/1623561968626941953?s=20&t=sp6Y-hSy... ; https://twitter.com/IntelCrab/status/1623562915331690500?s=20&t=sp6Y-hSy... ; https://twitter.com/IntelCrab/status/1623563711859290113?s=20&t=sp6Y-hSy...;\xa0https://t.me/zoda_gov_ua/16537\xa0;\xa0https://t.me/skarlatop/976\xa0;\xa0https:...\n\n[38]\xa0https://t.me/Yevtushenko_E/2427\xa0;\xa0https://www.facebook.com/GeneralStaf...\n\n[39]\xa0https://t.me/grey_zone/17129\n\n[40]\xa0https://www.understandingwar.org/backgrounder/russian-offensive-campaign...\n\n[41]\xa0https://59 dot ru /text/gorod/2023/02/07/72040580/?utm_source=telegram&utm_medium=messenger&utm_campaign=59\n\n[42]\xa0https://59 dot ru /text/gorod/2023/02/07/72040580/?utm_source=telegram&utm_medium=messenger&utm_campaign=59\n\n[43]\xa0https://www.politnavigator dot net/my-vse-budem-udivleny-kogda-nachnjotsya-nastuplenie-rogov.html\n\n[44]\xa0https://isw.pub/UkrWar020223\n\n[45]\xa0http://kremlin\xa0dot ru/events/president/news/70482\n\n[46]\xa0http://kre

```
mlin\xa0dot ru/events/president/news/70482\n\n[47]\xa0https://t.me/push
ilindenis/3166\n\n[48]\xa0https://t.me/BalitskyEV/778\xa0;\n\n[49]\xa0h
ttps://t.me/BalitskyEV/778\n\n[50]\xa0https://isw.pub/UkrWar011323\n\n
[51]\xa0http://kremlin\xa0dot ru/events/president/news/70482\n\n[52]\xa
0http://kremlin\xa0dot ru/events/president/news/70482\n\n[53]\xa0http
s://t.me/vrogov/7593;\xa0https://t.me/vrogov/7582\xa0;\xa0https://t.me/
rea...\n\n[54]\xa0https://t.me/vrogov/7593\xa0;\xa0https://t.me/rybar/4
3417\n\n[55]\xa0https://www.understandingwar.org/backgrounder/russian-o
ffensive-campaign...\n\n[56]\xa0https://t.me/modmilby/22518\n\n[57]\xa0
https://t.me/modmilby/22515\n\n[58]\xa0https://t.me/grey_zone/17123\n\n
File Attachments:\n\nZaporizhia Battle Map Draft February 09,2023.png\n
\nKherson-Mykolaiv Battle Map Draft February 09,2023.png\n\nDonetsk Bat
tle Map Draft February 09,2023.png\n\nKharkiv Battle Map Draft February
09,2023.png\n\nDraftUkraineCoTFebruary09,2023.png\n\n1400 16th Street N
W, Suite 515 Washington, DC 20036', metadata={'source': 'https://www.un
derstandingwar.org/backgrounder/russian-offensive-campaign-assessment-f
ebruary-9-2023'})]
```

Selenium Loader

This covers how to load HTML documents from a list of URLs using the SeleniumURLLoader.

Using selenium allows us to load pages that require JavaScript to render.

In [ ]:
```
!pip install selenium
```

In [ ]:
```
!/Users/cherifbenham/.pyenv/versions/3.9.16/bin/python3.9 -m pip install
```

In [ ]:
```
from langchain.document_loaders import SeleniumURLLoader
```

In [204…
```
urls = [
    "https://www.youtube.com/watch?v=dQw4w9WgXcQ",
    "https://goo.gl/maps/NDSHwePEyaHMFGwh8"
]
```

In [ ]:
```
loader=SeleniumURLLoader(urls=urls)
docs=loader.load()
docs
```

Web Loader

This covers how to load all text from webpages into a document format that we can use downstream. For more custom logic for loading webpages look at some child class examples such as IMSDbLoader, AZLyricsLoader, and CollegeConfidentialLoader

In [ ]:
```
from langchain.document_loaders import WebBaseLoader
```

In [206…
```
loader=WebBaseLoader("https://www.festival-cannes.com/presse/communiques/
```

In [ ]:
```
docs=loader.load()
docs
```

Concurrent Scraping

You can speed up the scraping process by scraping and parsing multiple urls concurrently.

There are reasonable limits to concurrent requests, defaulting to 2 per second. If you aren't concerned about being a good citizen, or you control the server you are scraping and don't care about load, you can change the requests_per_second parameter to increase the max concurrent requests. Note, while this will speed up the scraping process, but may cause the server to block you. Be careful!

In [208…
```
!pip install nest_asyncio

# fixes a bug with asyncio and jupyter
import nest_asyncio

nest_asyncio.apply()
```

```
Requirement already satisfied: nest_asyncio in /Users/cherifbenham/.pyenv/
versions/3.9.16/lib/python3.9/site-packages (1.5.6)
```

In [ ]:
```
loader = WebBaseLoader(["https://www.festival-cannes.com/presse/communiqu
                        , "https://google.com"])
loader.requests_per_second = 1
docs = loader.aload()
docs
```

XML Loading

In [ ]:
```
loader = WebBaseLoader("https://www.govinfo.gov/content/pkg/CFR-2018-titl
loader.default_parser = "xml"
docs = loader.load()
docs
```

WhatsApp Chat

In [211…
```
from langchain.document_loaders import WhatsAppChatLoader
```

Word Documents

In [ ]:
```
from langchain.document_loaders import UnstructuredWordDocumentLoader
```

CSV Agents

In [ ]:
```
from langchain.document_loaders.csv_loader import CSVLoader
loader = CSVLoader(file_path='../input/all_films.csv', source_column="Eng
```

```
data = loader.load()
print(data)

from langchain.agents import create_csv_agent
from langchain.llms import OpenAI
agent = create_csv_agent(OpenAI(temperature=0), '../input/all_films.csv',
agent.run("how are the film directors that will be present in festival de
```

Name: Director(s), dtype: object I now know the directors that will be present in festival de cannes 2

Final Answer: The directors that will be present in festival de cannes 2023 are Nuri Bilge Ceylan, Justine Triet, Wes Anderson, Ramata-Toulaye Sy, Alice Rohrwacher, Jessica Hausner, Aki Kaurismäki, Karim Aïnouz, Kaouther Ben Hania, Nanni Moretti, Marco Bellocchio, Catherine Breillat, Todd Haynes, Hirokazu Kore-eda, Ken Loach, Wim Wenders, Tran Anh Hung, Wang Bing, Jonathan Glazer, Rodrigo Moreno, Molly Manning Walker, Mohamed Kordofani, Joao Salaviza, Renée Nader Messora, Monia Chokri, Asmae El Moudir, Felipe Gálvez, Baloji Tshiani, Anthony Chen, Stéphanie Di Giusto, Warwick Thornton, Zoljargal Purevdash, Kim Chang-Hoon, Ali Asgari, Alireza Khatami, Delphine Deloget, Kamal Lazraq, Thomas Cailley, Kim Jee-woon,

> Finished chain.

'The directors that will be present in festival de cannes 2023 are Nuri Bilge Ceylan, Justine Triet, Wes Anderson,

Ramata-Toulaye Sy, Alice Rohrwacher, Jessica Hausner, Aki Kaurismäki, Karim Aïnouz, Kaouther Ben Hania, Nanni Moretti, Marco Bellocchio, Catherine Breillat, Todd Haynes, Hirokazu Kore-eda, Ken Loach, Wim Wenders, Tran Anh Hung, Wang Bing, Jonathan Glazer, Rodrigo Moreno, Molly Manning Walker, Mohamed Kordofani, Joao Salaviza, Renée Nader Messora, Monia Chokri, Asmae El Moudir, Felipe Gálvez, Baloji Tshiani, Anthony Chen, Stéphanie Di Giusto, Warwick Thornton, Zoljargal Purevdash, Kim Chang-Hoon, Ali Asgari, Alireza Khatami, Delphine Deloget, Kamal Lazraq, Thomas Cailley, Kim Jee-woon,'

Text splitters

When you want to deal with long pieces of text, it is necessary to split up that text into chunks. As simple as this sounds, there is a lot of potential complexity here. Ideally, you want to keep the semantically related pieces of text together. What "semantically related" means could depend on the type of text. This notebook showcases several ways to do that.

At a high level, text splitters work as following:

- Split the text up into small, semantically meaningful chunks (often sentences).
- Start combining these small chunks into a larger chunk until you reach a certain size (as measured by some function).
- Once you reach that size, make that chunk its own piece of text and then start creating a new chunk of text with some overlap (to keep context between chunks).

That means there two different axes along which you can customize your text splitter:

- How the text is split
- How the chunk size is measured

Text splitters that are supported

- Character Text Splitter
- Hugging Face Length Function
- Latex Text Splitter
- Markdown Text Splitter
- NLTK Text Splitter
- Python Code Text Splitter
- RecursiveCharacterTextSplitter
- Spacy Text Splitter
- tiktoken (OpenAI) Length Function
- TiktokenText Splitter

Recommended: RecursiveCharacterTextSplitter

The default recommended text splitter is the RecursiveCharacterTextSplitter. This text splitter takes a list of characters. It tries to create chunks based on splitting on the first character, but if any chunks are too large it then moves onto the next character, and so forth. By default the characters it tries to split on are ["\n\n", "\n", " ", ""]

In addition to controlling which characters you can split on, you can also control a few other things:

- length_function: how the length of chunks is calculated. Defaults to just counting number of characters, but it's pretty common to pass a token counter here.
- chunk_size: the maximum size of your chunks (as measured by the length function).
- chunk_overlap: the maximum overlap between chunks. It can be nice to have some overlap to maintain some continuity between chunks (eg do a sliding window).

In [212…

```
# This is a long document we can split up.
with open('../input/state_of_the_union.txt') as f:
    state_of_the_union = f.read()

from langchain.text_splitter import RecursiveCharacterTextSplitter

text_splitter = RecursiveCharacterTextSplitter(
    # Set a really small chunk size, just to show.
    chunk_size = 200,
    chunk_overlap  = 20,
    length_function = len,
)

texts = text_splitter.create_documents([state_of_the_union])
print(texts[0])
print(texts[1])
```

```
page_content='Madam Speaker, Madam Vice President, our First Lady and Seco
nd Gentleman. Members of Congress and the Cabinet. Justices of the Supreme
Court. My fellow Americans.' metadata={}
page_content='Last year COVID-19 kept us apart. This year we are finally t
ogether again. \n\nTonight, we meet as Democrats Republicans and Independe
nts. But most importantly as Americans.' metadata={}
```

OpenAI Text Splitter - tiktoken

You can also use tiktoken, a open source tokenizer package from OpenAI to estimate tokens used. Will probably be more accurate for their models.

In [213…
```
with open("../input/state_of_the_union.txt") as f:
    state_of_the_union=f.read()

from langchain.text_splitter import CharacterTextSplitter
text_splitter = CharacterTextSplitter.from_tiktoken_encoder(chunk_size=10

texts=text_splitter.split_text(state_of_the_union)
print(texts[0])
```

```
Madam Speaker, Madam Vice President, our First Lady and Second Gentleman.
Members of Congress and the Cabinet. Justices of the Supreme Court. My fel
low Americans.

Last year COVID-19 kept us apart. This year we are finally together again.

Tonight, we meet as Democrats Republicans and Independents. But most impor
tantly as Americans.

With a duty to one another to the American people to the Constitution.
```

Vectorstores

Vectorstores are one of the most important components of building indexes.

Vectorestore creation

In [214…
```
with open("../input/state_of_the_union.txt") as f:
    state_of_the_union = f.read()
```

In [215…
```
from langchain.embeddings.openai import OpenAIEmbeddings
embeddings=OpenAIEmbeddings()
```

In [216…
```
from langchain.vectorstores import Chroma
```

In [217…
```
from langchain.text_splitter import CharacterTextSplitter
text_splitter=CharacterTextSplitter(chunk_size=1000, chunk_overlap=0)
texts=text_splitter.split_text(state_of_the_union)
```

In [218…
```
docsearch=Chroma.from_texts(texts,embeddings)
```

```
Using embedded DuckDB without persistence: data will be transient
```

Vectorestore Search similarity

In [219…
```
query='what did the president say about ecology'
```

In [220…
```
docsearch.similarity_search(query)
```

Out[220]:
```
[Document(page_content='It is going to transform America and put us on a path to win the economic competition of the 21st Century that we face with the rest of the world—particularly with China.  \n\nAs I've told Xi Jinping, it is never a good bet to bet against the American people. \n\nWe'll create good jobs for millions of Americans, modernizing roads, airports, ports, and waterways all across America. \n\nAnd we'll do it all to withstand the devastating effects of the climate crisis and promote environmental justice. \n\nWe'll build a national network of 500,000 electric vehicle charging stations, begin to replace poisonous lead pipes—so every child—and every American—has clean water to drink at home and at school, provide affordable high-speed internet for every American—urban, suburban, rural, and tribal communities. \n\n4,000 projects have already been announced. \n\nAnd tonight, I'm announcing that this year we will start fixing over 65,000 miles of highway and 1,500 bridges in disrepair.', metadata={}),
 Document(page_content='Vice President Harris and I ran for office with a new economic vision for America. \n\nInvest in America. Educate Americans. Grow the workforce. Build the economy from the bottom up  \nand the middle out, not from the top down.  \n\nBecause we know that when the middle class grows, the poor have a ladder up and the wealthy do very well. \n\nAmerica used to have the best roads, bridges, and airports on Earth. \n\nNow our infrastructure is ranked 13th in the world. \n\nWe won't be able to compete for the jobs of the 21st Century if we don't fix that. \n\nThat's why it was so important to pass the Bipartisan Infrastructure Law—the most sweeping investment to rebuild America in history. \n\nThis was a bipartisan effort, and I want to thank the members of both parties who worked to make it happen. \n\nWe're done talking about infrastructure weeks. \n\nWe're going to have an infrastructure decade.', metadata={}),
 Document(page_content='We got more than 130 countries to agree on a global minimum tax rate so companies can't get out of paying their taxes at home by shipping jobs and factories overseas. \n\nThat's why I've proposed closing loopholes so the very wealthy don't pay a lower tax rate than a teacher or a firefighter.  \n\nSo that's my plan. It will grow the economy and lower costs for families. \n\nSo what are we waiting for? Let's get this done. And while you're at it, confirm my nominees to the Federal Reserve, which plays a critical role in fighting inflation.  \n\nMy plan will not only lower costs to give families a fair shot, it will lower the deficit. \n\nThe previous Administration not only ballooned the deficit with tax cuts for the very wealthy and corporations, it undermined the watchdogs whose job was to keep pandemic relief funds from being wasted. \n\nBut in my administration, the watchdogs have been welcomed back.', metadata={}),
 Document(page_content='And for our LGBTQ+ Americans, let's finally get the bipartisan Equality Act to my desk. The onslaught of state laws targeting transgender Americans and their families is wrong. \n\nAs I said last year, especially to our younger transgender Americans, I will always have your back as your President, so you can be yourself and reach your God-given potential. \n\nWhile it often appears that we never agree, that isn't true. I signed 80 bipartisan bills into law last year. From preventing government shutdowns to protecting Asian-Americans from still-too-common hate crimes to reforming military justice. \n\nAnd soon, we'll strengthen the Violence Against Women Act that I first wrote three decades ago. It is important for us to show the nation that we can come together and do big things. \n\nSo tonight I'm offering a Unity Agenda for the Nation. Four big things we can do together.  \n\nFirst, beat the opioid epidemic.', metadata={})]
```

In [221…
```
docs=docsearch.similarity_search(query)
```

```
docs[0].page_content
```

Out[221]:
```
'It is going to transform America and put us on a path to win the econom
ic competition of the 21st Century that we face with the rest of the wor
ld—particularly with China.  \n\nAs I've told Xi Jinping, it is never a
good bet to bet against the American people. \n\nWe'll create good jobs
for millions of Americans, modernizing roads, airports, ports, and water
ways all across America. \n\nAnd we'll do it all to withstand the devast
ating effects of the climate crisis and promote environmental justice.
\n\nWe'll build a national network of 500,000 electric vehicle charging
stations, begin to replace poisonous lead pipes—so every child—and every
American—has clean water to drink at home and at school, provide afforda
ble high-speed internet for every American—urban, suburban, rural, and t
ribal communities. \n\n4,000 projects have already been announced. \n\nA
nd tonight, I'm announcing that this year we will start fixing over 65,0
00 miles of highway and 1,500 bridges in disrepair.'
```

Adding a text to a vectorstore

In [222…
```
docsearch.add_texts(["spotify invests in ecology", "elon musk want to com
```

Out[222]:
```
['9ccf3ff0-e039-11ed-868d-0aed9a9c3245',
 '9ccf4158-e039-11ed-868d-0aed9a9c3245']
```

In [223…
```
query="what's the plan of elon musk"

docsearch.similarity_search(query)
```

Out[223]:
```
[Document(page_content='elon musk want to compete with openai', metadata
={}),
 Document(page_content='I have a better plan to fight inflation. \n\nLow
er your costs, not your wages. \n\nMake more cars and semiconductors in
America. \n\nMore infrastructure and innovation in America. \n\nMore goo
ds moving faster and cheaper in America. \n\nMore jobs where you can ear
n a good living in America. \n\nAnd instead of relying on foreign supply
chains, let's make it in America. \n\nEconomists call it "increasing the
productive capacity of our economy." \n\nI call it building a better Ame
rica. \n\nMy plan to fight inflation will lower your costs and lower the
deficit. \n\n17 Nobel laureates in economics say my plan will ease long-
term inflationary pressures. Top business leaders and most Americans sup
port my plan. And here's the plan: \n\nFirst — cut the cost of prescript
ion drugs. Just look at insulin. One in ten Americans has diabetes. In V
irginia, I met a 13-year-old boy named Joshua Davis.', metadata={}),
 Document(page_content='It is going to transform America and put us on a
path to win the economic competition of the 21st Century that we face wi
th the rest of the world—particularly with China.  \n\nAs I've told Xi J
inping, it is never a good bet to bet against the American people. \n\nW
e'll create good jobs for millions of Americans, modernizing roads, airp
orts, ports, and waterways all across America. \n\nAnd we'll do it all t
o withstand the devastating effects of the climate crisis and promote en
vironmental justice. \n\nWe'll build a national network of 500,000 elect
ric vehicle charging stations, begin to replace poisonous lead pipes—so
every child—and every American—has clean water to drink at home and at s
chool, provide affordable high-speed internet for every American—urban,
suburban, rural, and tribal communities. \n\n4,000 projects have already
been announced. \n\nAnd tonight, I'm announcing that this year we will s
tart fixing over 65,000 miles of highway and 1,500 bridges in disrepai
r.', metadata={}),
 Document(page_content='spotify invests in ecology', metadata={})]
```

Create vectorstore from documents

handy when the documents have metadata

In [224…
```
state_of_the_union_2=state_of_the_union
```

In [225…
```
documents=text_splitter.create_documents(
    [state_of_the_union,
     state_of_the_union_2
    ],
    metadatas =
    [
        {'source':"state_of_the_union",
        'date_of_publication':"2023-04-19"},

        {'source':"state_of_the_union_2",
        'date_of_publication':"2023-04-21"}
    ])
```

In [226…
```
documents[0]
```

Out[226]:
```
Document(page_content='Madam Speaker, Madam Vice President, our First La
dy and Second Gentleman. Members of Congress and the Cabinet. Justices o
f the Supreme Court. My fellow Americans.  \n\nLast year COVID-19 kept u
s apart. This year we are finally together again. \n\nTonight, we meet a
s Democrats Republicans and Independents. But most importantly as Americ
ans. \n\nWith a duty to one another to the American people to the Consti
tution. \n\nAnd with an unwavering resolve that freedom will always triu
mph over tyranny. \n\nSix days ago, Russia's Vladimir Putin sought to sh
ake the foundations of the free world thinking he could make it bend to
his menacing ways. But he badly miscalculated. \n\nHe thought he could r
oll into Ukraine and the world would roll over. Instead he met a wall of
strength he never imagined. \n\nHe met the Ukrainian people. \n\nFrom Pr
esident Zelenskyy to every Ukrainian, their fearlessness, their courage,
their determination, inspires the world.', metadata={'source': 'state_of
_the_union', 'date_of_publication': '2023-04-19'})
```

In [227…

```
docsearch=Chroma.from_documents(documents,embeddings)
docsearch.similarity_search("ecology")
```

```
Using embedded DuckDB without persistence: data will be transient
```

Out[227]: [Document(page_content='We can do all this while keeping lit the torch o
f liberty that has led generations of immigrants to this land—my forefat
hers and so many of yours. \n\nProvide a pathway to citizenship for Drea
mers, those on temporary status, farm workers, and essential workers. \n
\nRevise our laws so businesses have the workers they need and families
don’t wait decades to reunite. \n\nIt’s not only the right thing to do—i
t’s the economically smart thing to do. \n\nThat’s why immigration refor
m is supported by everyone from labor unions to religious leaders to the
U.S. Chamber of Commerce. \n\nLet’s get it done once and for all. \n\nAd
vancing liberty and justice also requires protecting the rights of wome
n. \n\nThe constitutional right affirmed in Roe v. Wade—standing precede
nt for half a century—is under attack as never before. \n\nIf we want to
go forward—not backward—we must protect access to health care. Preserve
a woman’s right to choose. And let’s continue to advance maternal health
care in America.', metadata={'source': 'state_of_the_union', 'date_of_pu
blication': '2023-04-19'}),
 Document(page_content='We can do all this while keeping lit the torch o
f liberty that has led generations of immigrants to this land—my forefat
hers and so many of yours. \n\nProvide a pathway to citizenship for Drea
mers, those on temporary status, farm workers, and essential workers. \n
\nRevise our laws so businesses have the workers they need and families
don’t wait decades to reunite. \n\nIt’s not only the right thing to do—i
t’s the economically smart thing to do. \n\nThat’s why immigration refor
m is supported by everyone from labor unions to religious leaders to the
U.S. Chamber of Commerce. \n\nLet’s get it done once and for all. \n\nAd
vancing liberty and justice also requires protecting the rights of wome
n. \n\nThe constitutional right affirmed in Roe v. Wade—standing precede
nt for half a century—is under attack as never before. \n\nIf we want to
go forward—not backward—we must protect access to health care. Preserve
a woman’s right to choose. And let’s continue to advance maternal health
care in America.', metadata={'source': 'state_of_the_union_2', 'date_of_
publication': '2023-04-21'}),
 Document(page_content='And built the strongest, freest, and most prospe
rous nation the world has ever known. \n\nNow is the hour. \n\nOur momen
t of responsibility. \n\nOur test of resolve and conscience, of history
itself. \n\nIt is in this moment that our character is formed. Our purpo
se is found. Our future is forged. \n\nWell I know this nation. \n\nWe
will meet the test. \n\nTo protect freedom and liberty, to expand fairne
ss and opportunity. \n\nWe will save democracy. \n\nAs hard as these tim
es have been, I am more optimistic about America today than I have been
my whole life. \n\nBecause I see the future that is within our grasp. \n
\nBecause I know there is simply nothing beyond our capacity. \n\nWe are
the only nation on Earth that has always turned every crisis we have fac
ed into an opportunity. \n\nThe only nation that can be defined by a sin
gle word: possibilities. \n\nSo on this night, in our 245th year as a na
tion, I have come to report on the State of the Union.', metadata={'sour
ce': 'state_of_the_union', 'date_of_publication': '2023-04-19'}),
 Document(page_content='And built the strongest, freest, and most prospe
rous nation the world has ever known. \n\nNow is the hour. \n\nOur momen
t of responsibility. \n\nOur test of resolve and conscience, of history
itself. \n\nIt is in this moment that our character is formed. Our purpo
se is found. Our future is forged. \n\nWell I know this nation. \n\nWe
will meet the test. \n\nTo protect freedom and liberty, to expand fairne
ss and opportunity. \n\nWe will save democracy. \n\nAs hard as these tim
es have been, I am more optimistic about America today than I have been
my whole life. \n\nBecause I see the future that is within our grasp. \n
\nBecause I know there is simply nothing beyond our capacity. \n\nWe are
the only nation on Earth that has always turned every crisis we have fac
ed into an opportunity. \n\nThe only nation that can be defined by a sin
gle word: possibilities. \n\nSo on this night, in our 245th year as a na

```
tion, I have come to report on the State of the Union.', metadata={'sour
ce': 'state_of_the_union_2', 'date_of_publication': '2023-04-21'})]
```

Search similarities with scores

In [228…

```
query = "What did the president say about jobs"
docs = docsearch.similarity_search_with_score(query)

docs
```

Out[228]: [(Document(page_content='It is going to transform America and put us on a path to win the economic competition of the 21st Century that we face with the rest of the world—particularly with China. \n\nAs I've told Xi Jinping, it is never a good bet to bet against the American people. \n\nWe'll create good jobs for millions of Americans, modernizing roads, airports, ports, and waterways all across America. \n\nAnd we'll do it all to withstand the devastating effects of the climate crisis and promote environmental justice. \n\nWe'll build a national network of 500,000 electric vehicle charging stations, begin to replace poisonous lead pipes—so every child—and every American—has clean water to drink at home and at school, provide affordable high-speed internet for every American—urban, suburban, rural, and tribal communities. \n\n4,000 projects have already been announced. \n\nAnd tonight, I'm announcing that this year we will start fixing over 65,000 miles of highway and 1,500 bridges in disrepair.', metadata={'source': 'state_of_the_union', 'date_of_publication': '2023-04-19'}),
  0.36300694942474365),
 (Document(page_content='It is going to transform America and put us on a path to win the economic competition of the 21st Century that we face with the rest of the world—particularly with China. \n\nAs I've told Xi Jinping, it is never a good bet to bet against the American people. \n\nWe'll create good jobs for millions of Americans, modernizing roads, airports, ports, and waterways all across America. \n\nAnd we'll do it all to withstand the devastating effects of the climate crisis and promote environmental justice. \n\nWe'll build a national network of 500,000 electric vehicle charging stations, begin to replace poisonous lead pipes—so every child—and every American—has clean water to drink at home and at school, provide affordable high-speed internet for every American—urban, suburban, rural, and tribal communities. \n\n4,000 projects have already been announced. \n\nAnd tonight, I'm announcing that this year we will start fixing over 65,000 miles of highway and 1,500 bridges in disrepair.', metadata={'source': 'state_of_the_union_2', 'date_of_publication': '2023-04-21'}),
  0.36300694942474365),
 (Document(page_content='And unlike the $2 Trillion tax cut passed in the previous administration that benefitted the top 1% of Americans, the American Rescue Plan helped working people—and left no one behind. \n\nAnd it worked. It created jobs. Lots of jobs. \n\nIn fact—our economy created over 6.5 Million new jobs just last year, more jobs created in one year \nthan ever before in the history of America. \n\nOur economy grew at a rate of 5.7% last year, the strongest growth in nearly 40 years, the first step in bringing fundamental change to an economy that hasn't worked for the working people of this nation for too long. \n\nFor the past 40 years we were told that if we gave tax breaks to those at the very top, the benefits would trickle down to everyone else. \n\nBut that trickle-down theory led to weaker economic growth, lower wages, bigger deficits, and the widest gap between those at the top and everyone else in nearly a century.', metadata={'source': 'state_of_the_union', 'date_of_publication': '2023-04-19'}),
  0.3709045648574829),
 (Document(page_content='And unlike the $2 Trillion tax cut passed in the previous administration that benefitted the top 1% of Americans, the American Rescue Plan helped working people—and left no one behind. \n\nAnd it worked. It created jobs. Lots of jobs. \n\nIn fact—our economy created over 6.5 Million new jobs just last year, more jobs created in one year \nthan ever before in the history of America. \n\nOur economy grew at a rate of 5.7% last year, the strongest growth in nearly 40 years, the first step in bringing fundamental change to an economy that hasn't worked for the working people of this nation for too long. \n\nFor the past 40 years we were told that if we gave tax breaks to those at the very

```
top, the benefits would trickle down to everyone else. \n\nBut that tric
kle-down theory led to weaker economic growth, lower wages, bigger defic
its, and the widest gap between those at the top and everyone else in ne
arly a century.', metadata={'source': 'state_of_the_union_2', 'date_of_p
ublication': '2023-04-21'}),
  0.3709045648574829)]
```

Initialize PeristedChromaDB

Create embeddings for each chunk and insert into the Chroma vector database. The persist_directory argument tells ChromaDB where to store the database when it's persisted.

In [229…
```
persist_directory='db'
embedding=OpenAIEmbeddings()

vectordb=Chroma.from_documents(documents=documents, embedding=embedding,
```

```
Using embedded DuckDB with persistence: data will be stored in: db
```

Persist the database

We should call persist() to ensure the embeddings are written to disk.

In [230…
```
vectordb.persist()
```

In [231…
```
!tree db
```

```
db
├── chroma-collections.parquet
├── chroma-embeddings.parquet
└── index
    ├── id_to_uuid_68f49e20-2968-46d0-8eb7-ed1ae4f2ae4c.pkl
    ├── index_68f49e20-2968-46d0-8eb7-ed1ae4f2ae4c.bin
    ├── index_metadata_68f49e20-2968-46d0-8eb7-ed1ae4f2ae4c.pkl
    └── uuid_to_id_68f49e20-2968-46d0-8eb7-ed1ae4f2ae4c.pkl

1 directory, 6 files
```

Load the database from disk

In [232…
```
vectordb=Chroma(persist_directory=persist_directory,embedding_function=em
```

```
Using embedded DuckDB with persistence: data will be stored in: db
```

Chroma as retriever

In [233…
```
retriever=vectordb.as_retriever(search_type="similarity")
```

In [234…
```
retriever.get_relevant_documents(query)
```

Out[234]: [Document(page_content='It is going to transform America and put us on a path to win the economic competition of the 21st Century that we face with the rest of the world—particularly with China. \n\nAs I've told Xi Jinping, it is never a good bet to bet against the American people. \n\nWe'll create good jobs for millions of Americans, modernizing roads, airports, ports, and waterways all across America. \n\nAnd we'll do it all to withstand the devastating effects of the climate crisis and promote environmental justice. \n\nWe'll build a national network of 500,000 electric vehicle charging stations, begin to replace poisonous lead pipes—so every child—and every American—has clean water to drink at home and at school, provide affordable high-speed internet for every American—urban, suburban, rural, and tribal communities. \n\n4,000 projects have already been announced. \n\nAnd tonight, I'm announcing that this year we will start fixing over 65,000 miles of highway and 1,500 bridges in disrepair.', metadata={'source': 'state_of_the_union', 'date_of_publication': '2023-04-19'}),
 Document(page_content='It is going to transform America and put us on a path to win the economic competition of the 21st Century that we face with the rest of the world—particularly with China. \n\nAs I've told Xi Jinping, it is never a good bet to bet against the American people. \n\nWe'll create good jobs for millions of Americans, modernizing roads, airports, ports, and waterways all across America. \n\nAnd we'll do it all to withstand the devastating effects of the climate crisis and promote environmental justice. \n\nWe'll build a national network of 500,000 electric vehicle charging stations, begin to replace poisonous lead pipes—so every child—and every American—has clean water to drink at home and at school, provide affordable high-speed internet for every American—urban, suburban, rural, and tribal communities. \n\n4,000 projects have already been announced. \n\nAnd tonight, I'm announcing that this year we will start fixing over 65,000 miles of highway and 1,500 bridges in disrepair.', metadata={'source': 'state_of_the_union_2', 'date_of_publication': '2023-04-21'}),
 Document(page_content='It is going to transform America and put us on a path to win the economic competition of the 21st Century that we face with the rest of the world—particularly with China. \n\nAs I've told Xi Jinping, it is never a good bet to bet against the American people. \n\nWe'll create good jobs for millions of Americans, modernizing roads, airports, ports, and waterways all across America. \n\nAnd we'll do it all to withstand the devastating effects of the climate crisis and promote environmental justice. \n\nWe'll build a national network of 500,000 electric vehicle charging stations, begin to replace poisonous lead pipes—so every child—and every American—has clean water to drink at home and at school, provide affordable high-speed internet for every American—urban, suburban, rural, and tribal communities. \n\n4,000 projects have already been announced. \n\nAnd tonight, I'm announcing that this year we will start fixing over 65,000 miles of highway and 1,500 bridges in disrepair.', metadata={'source': 'state_of_the_union', 'date_of_publication': '2023-04-19'}),
 Document(page_content='It is going to transform America and put us on a path to win the economic competition of the 21st Century that we face with the rest of the world—particularly with China. \n\nAs I've told Xi Jinping, it is never a good bet to bet against the American people. \n\nWe'll create good jobs for millions of Americans, modernizing roads, airports, ports, and waterways all across America. \n\nAnd we'll do it all to withstand the devastating effects of the climate crisis and promote environmental justice. \n\nWe'll build a national network of 500,000 electric vehicle charging stations, begin to replace poisonous lead pipes—so every child—and every American—has clean water to drink at home and at school, provide affordable high-speed internet for every American—urban, suburban, rural, and tribal communities. \n\n4,000 projects have already

been announced. \n\nAnd tonight, I’m announcing that this year we will s
tart fixing over 65,000 miles of highway and 1,500 bridges in disrepai
r.', metadata={'source': 'state_of_the_union_2', 'date_of_publication':
'2023-04-21'})]