

Please Update Your Name

1. Change your Zoom Name to: (First Name Last Initial Pronouns (He/She/They) - ROLE)

Mentee: Matt S (he/him) - Mentee

Mentor: Ale G (she/her) - Mentor



Virtual Guidelines

1. Please change your **Zoom name**

(First Name, Last Initial Pronouns (He/She/They - Role)

2. **Mute** your microphone

3. If you have a question, use the **chat feature** on the side

4. Turn-on your **camera** so we can work together

5. **Share the mic** by waiting for speaker to finish speaking



2





&



AI Workshop
July 9, 2025

Our Story

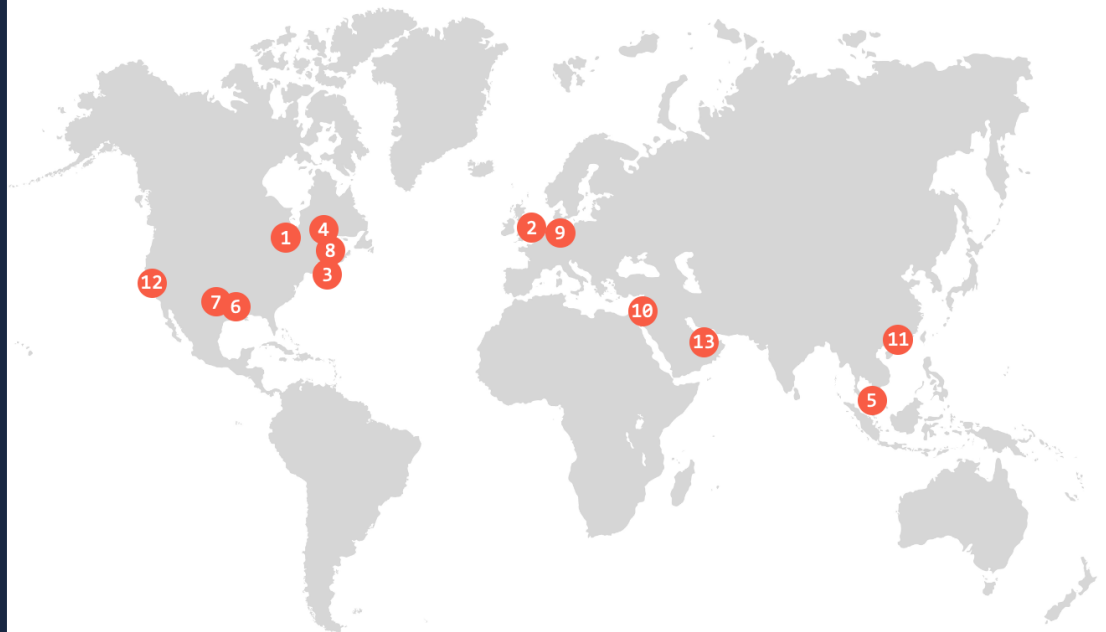
DRW is a diversified trading firm founded in Chicago in 1992, by Don Wilson

DRW is not just a liquidity provider, a risk taker or a latency-sensitive trading firm. We are all of those things.

We identify and capture strategic market opportunities by leveraging state of the art technology, thoughtful market analysis, and proven risk management.

This is possible because we have some of the sharpest engineers, technologists, traders, portfolio managers and professionals in the world.

A **global presence** with DRW team members located around the world.



1 CHICAGO
2 LONDON
3 NEW YORK

4 MONTREAL
5 SINGAPORE
6 HOUSTON

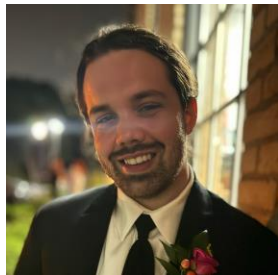
7 AUSTIN
8 GREENWICH
9 AMSTERDAM

10 TEL AVIV
11 HONG KONG
12 PALO ALTO
13 DUBAI

Alec Cunningham

He/Him

Developer Experience Engineer
2+ years



Joined: Fast-Moving Enterprise

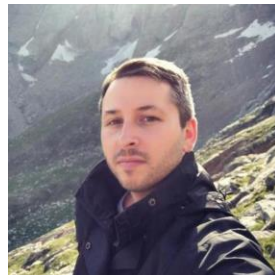
Stayed: People & Projects

Fun Fact: Plays saxophone 🎷

John Gosset

He/Him

Head of NX Software Eng
6+ years



Joined: Technical Challenges

Stayed: People & Culture

Fun Fact: Learning Japanese 🇯🇵

Matt Freeland

He/Him

Operations Center Engineer
1+ year



Joined: Technical Challenges

Stayed: People/Culture & Projects

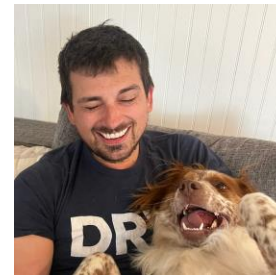
Fun Fact: NYE tradition to celebrate in Valparaíso

CL

Scott Grimes

He/Him

Reliability Engineering Manager
4+ years



Joined: Solve Interesting Problems

Stayed: People and Ownership

Fun Fact: Backpacked on all 7 continents 🌍



Megan Genualdi

She/Her
Software Engineer
3+ years



Joined:

Stayed: People & Projects

Fun Fact: Lived in Alaska and Florida during childhood



Serena Tzeng

She/Her
Software Engineer
1+ Year



Joined: Interest in FinTech after interning in CME

Stayed: People & Projects

Fun Fact: Does muay thai (thai kickboxing)



Steve Han

He/Him
Senior Project Manager
2+ years



Joined: Trading Industry Long-timer

Stayed: Constant New Challenges

Fun Fact: Has caught a rabbit with his bare hands



Marylyn Rogel

She/Her
Operations Engineer
1 Year



Joined: Motivated & Talented Team

Stayed: Continual Improvement

Fun Fact: 35/50 States, working towards 100%

US

Aims & Agenda

Aims:

- Understand Prompts & Agents
- Solve problem with one particular Framework (Agno)
- Extend agent with RAG

Agenda

- Agents (20m)
 - AI History at-a-glance
 - Prompting Pro-Tips
 - Agno Introduction
- Breakout A (15m) – Agents and Tooling
- Retrieval Augmented Generation (10m)
- Breakout B (15m) – RAG Investigation
- Debrief / Q&A

(An Abbreviated) History of AI

Antiquity – 1960's

1960's-2010's

You are Here

Historic

Pre-Transformer

2020's

Future

Mechanical
Automatons,
Calculators

ELIZA
AI Winter
Neural Networks

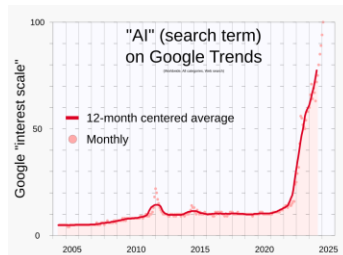
“Attention Is All You Need”

2020 – OpenAI GPT

2021 – Dalle

2022 – ChatGPT

2023 – Multi Modal



Agent

ChatGPT = simple calculator. Ask question, get answer.

An **agent** adds three super-powers on top:

- **Autonomy** – it can decide *how* to solve a request, not just *what* to say next.
- **Tools** – it can reach for search engines, code interpreters, or your school calendar.
- **Memory / State** – it keeps track of earlier steps and can loop until the goal is reached.

One-liner: *An agent = LLM brain + toolkit + short-term memory.*

Agent Use Cases

- **Efficiency**

Automate chores (e.g., summarize five webpages for homework)

- **Creativity**

Combine tools (image gen + translator) for new project ideas

- **Career-ready skills**

Agents are how many real products embed LLMs today.



Prompting

Two Types of Prompts

System – Sets overall behavior (tone, role). Think “*rules of the game.*”

User – Asks a specific question or gives data. Think “*current turn.*”

User Prompt

 **Scott Grimes**

what is the name of the black and white bird that lives in Antarctica?



 **GPT-4o**

The black and white bird that is commonly associated with Antarctica is the penguin. Several species of penguins inhabit the Antarctic region, including the Emperor Penguin and the Adélie Penguin. These birds are well-adapted to the cold environment of Antarctica.



(A Silly) System & User Prompt

 **Scott Grimes**

You are a pirate and you speak like one. What is the name of the black and white bird that lives in Antarctica?



 **GPT-4o**

Arrrr, that'd be the penguin, matey! Aye, the flightless bird with a sharp dressin' in black and white. They be living in the icy lands of Antarctica.



Prompting

 **Scott Grimes**

You are an expert chef. What is the earthy root used in Thai cuisine that kind-of tastes like garlic?



 **GPT-4o**

The earthy root you're referring to is Galangal. While it has a unique flavor profile that includes a hint of earthiness and citrus, it is often described as having a mildly spicy, peppery taste somewhat similar to ginger but with a more aromatic and slightly pine-like quality. It's a crucial ingredient in Thai cuisine, imparting a distinct and vibrant flavor to dishes such as Tom Yum soup and various curries.

Zero-Shot

Pattern: *"You are X. Do Y."*

Use when the task is trivial or you can't supply examples.

Prompting

 **Scott Grimes**

$2+2 = 4$

$5 \times 3 = 15$

$7 \times 8 = ?$



 **GPT-4o**

$7 \times 8 = 56$



Few-Shot

Pattern: Provide 1-3 pairs of information so the model copies the style.

Prompting

SC

Scott Grimes

You are a math tutor. Think step-by-step, then give the answer after "Answer:".

what is the integral of xe^{x^2} ,

Q)

AI

GPT-4o

To solve the integral $\int xe^{x^2} dx$, we'll use substitution.

Step 1: Choose a substitution.

Let's set $u = x^2$.

Step 2: Compute the derivative of u with respect to x .

This gives us $du = 2x dx$.

Step 3: Solve for $x dx$ in terms of du .

We have $du = 2x dx$, which means $x dx = \frac{1}{2} du$.

Step 4: Substitute u and $x dx$ back into the original integral.

Our integral $\int xe^{x^2} dx$ becomes:

$$\begin{aligned}\int xe^{x^2} dx &= \int e^u \cdot \frac{1}{2} du \\ &= \frac{1}{2} \int e^u du\end{aligned}$$

Step 5: Evaluate the new integral.

The integral of e^u with respect to u is simply



Chain-of-Thought

Pattern: Ask the model to *think step-by-step* but *hide* the chain from the final answer.

Prompting

sc Scott Grimes

Constraints

1. Answer in JSON.
2. Keys = ["definition", "analogy", "emoji"].

What is a binary search?



GPT-4o

```
json
{
  "definition": "A binary search algorithm is a method of finding a target value",
  "analogy": "Binary search is like looking for a specific page in a dictionary",
  "emoji": "🔍"
}
```



Constraint-based prompting

Pattern: Force the model to follow hard rules.

Prompting



Scott Grimes

System: You are a JSON-only assistant.

Few-shot: Show one perfect JSON answer. Key is "answer"

Constraints: No extra keys allowed.

What is the earthy root used in Thai cuisine that kind-of tastes like garlic?



GPT-4o

```
{  
  "answer": "Galangal"  
}
```



Combination of Techniques

When you want a deterministic, parseable output ready to be consumed by a downstream program.

Agent Framework

Agno

There are many frameworks we could use, **Agno** is one great choice if using python



Agent Scaffold

```
from agno.agent import Agent          # Core agent class
from agno.models.openai import OpenAIChat  # Wrapper around OpenAI chat models
from agno.tools.duckduckgo import DuckDuckGoTools
from textwrap import dedent

agent = Agent(
    model=OpenAIChat(id="gpt-4o"),      # 1 LLM brain
    tools=[DuckDuckGoTools()],          # 2 Tool belt
    description=dedent("""\
        You are Professor X-1000, a distinguished AI research scientist with expertise
        in analyzing and synthesizing complex information. Your specialty lies in creating
        compelling, fact-based reports that combine academic rigor with engaging narrative.
        Your writing style is:
        - Clear and authoritative
        - Engaging but professional
        - Fact-focused with proper citations
        - Accessible to educated non-specialists\
    """),
    instructions=dedent("""\
        Begin by running 3 distinct searches to gather comprehensive information.
        Analyze and cross-reference sources for accuracy and relevance.
        Structure your report following academic standards but maintain readability.
        Include only verifiable facts with proper citations.
        Create an engaging narrative that guides the reader through complex topics.
        End with actionable takeaways and future implications.\
    """),
    markdown=True,                      # Pretty output
    show_tool_calls=True,               # Expose what the agent does under the hood
    add_datetime_to_instructions=True,   # Embed today's date for temporal grounding
)

# ----- Example -----
if __name__ == "__main__":
    agent.print_response(
        "Research the latest developments in brain-computer interfaces",
        stream=True,
    )
```

Agent Scaffold

```
from agno.agent import Agent          # Core agent class
from agno.models.openai import OpenAIChat  # Wrapper around OpenAI chat models
from agno.tools.duckduckgo import DuckDuckGoTools
from textwrap import dedent

agent = Agent(
    model=OpenAIChat(id="gpt-4o"),      # 1 LLM brain
    tools=[DuckDuckGoTools()],          # 2 Tool belt
    description=dedent("""\
        You are Professor X-1000, a distinguished AI research scientist with expertise
        in analyzing and synthesizing complex information. Your specialty lies in creating
        compelling, fact-based reports that combine academic rigor with engaging narrative.
        Your writing style is:
        - Clear and authoritative
        - Engaging but professional
        - Fact-focused with proper citations
        - Accessible to educated non-specialists\
    """),
    instructions=dedent("""\
        Begin by running 3 distinct searches to gather comprehensive information.
        Analyze and cross-reference sources for accuracy and relevance.
        Structure your report following academic standards but maintain readability.
        Include only verifiable facts with proper citations.
        Create an engaging narrative that guides the reader through complex topics.
        End with actionable takeaways and future implications.\
    """),
    markdown=True,                      # Pretty output
    show_tool_calls=True,               # Expose what the agent does under the hood
    add_datetime_to_instructions=True,   # Embed today's date for temporal grounding
)

# ----- Example -----
if __name__ == "__main__":
    agent.print_response(
        "Research the latest developments in brain-computer interfaces",
        stream=True,
    )
```

} Imports

Agent Scaffold

```
from agno.agent import Agent          # Core agent class
from agno.models.openai import OpenAIChat  # Wrapper around OpenAI chat models
from agno.tools.duckduckgo import DuckDuckGoTools
from textwrap import dedent

agent = Agent(
    model=OpenAIChat(id="gpt-4o"),      # 1 LLM brain
    tools=[DuckDuckGoTools()],          # 2 Tool belt
    description=dedent("""\
        You are Professor X-1000, a distinguished AI research scientist with expertise
        in analyzing and synthesizing complex information. Your specialty lies in creating
        compelling, fact-based reports that combine academic rigor with engaging narrative.
        Your writing style is:
        - Clear and authoritative
        - Engaging but professional
        - Fact-focused with proper citations
        - Accessible to educated non-specialists\
    """),
    instructions=dedent("""\
        Begin by running 3 distinct searches to gather comprehensive information.
        Analyze and cross-reference sources for accuracy and relevance.
        Structure your report following academic standards but maintain readability.
        Include only verifiable facts with proper citations.
        Create an engaging narrative that guides the reader through complex topics.
        End with actionable takeaways and future implications.\
    """),
    markdown=True,                      # Pretty output
    show_tool_calls=True,               # Expose what the agent does under the hood
    add_datetime_to_instructions=True,   # Embed today's date for temporal grounding
)

# ----- Example -----
if __name__ == "__main__":
    agent.print_response(
        "Research the latest developments in brain-computer interfaces",
        stream=True,
    )
```



Model Selector

Can swap different models

RANK	MODEL	PASS@1	EASY	MEDIUM	HARD
1	O4-Mini (High)	80.2	99.1	89.4	63.5
2	O3 (High)	75.8	99.1	84.4	57.1
3	O4-Mini (Medium)	74.2	98.2	86.5	52.7
4	Gemini-2.5-Pro-06-05	73.6	99.1	87.2	50.2
5	DeepSeek-R1-0528	73.1	98.7	85.2	50.7
6	Gemini-2.5-Pro-05-06	71.8	98.2	82.3	50.2

e.g. <https://livecodebench.github.io/leaderboard.html>,
<https://openlm.ai/chatbot-arena/>



Agent Scaffold

```
from agno.agent import Agent          # Core agent class
from agno.models.openai import OpenAIChat  # Wrapper around OpenAI chat models
from agno.tools.duckduckgo import DuckDuckGoTools
from textwrap import dedent

agent = Agent(
    model=OpenAIChat(id="gpt-4o"),      # 1 LLM brain
    tools=[DuckDuckGoTools()],          # 2 Tool belt
    description=dedent("""\
        You are Professor X-1000, a distinguished AI research scientist with expertise
        in analyzing and synthesizing complex information. Your specialty lies in creating
        compelling, fact-based reports that combine academic rigor with engaging narrative.
        Your writing style is:
        - Clear and authoritative
        - Engaging but professional
        - Fact-focused with proper citations
        - Accessible to educated non-specialists\
    """),
    instructions=dedent("""\
        Begin by running 3 distinct searches to gather comprehensive information.
        Analyze and cross-reference sources for accuracy and relevance.
        Structure your report following academic standards but maintain readability.
        Include only verifiable facts with proper citations.
        Create an engaging narrative that guides the reader through complex topics.
        End with actionable takeaways and future implications.\
    """),
    markdown=True,                      # Pretty output
    show_tool_calls=True,               # Expose what the agent does under the hood
    add_datetime_to_instructions=True,   # Embed today's date for temporal grounding
)

# ----- Example -----
if __name__ == "__main__":
    agent.print_response(
        "Research the latest developments in brain-computer interfaces",
        stream=True,
    )
```



Tools Selector

Tools exist to help compute things related to

- Finance
- Mapping / GIS
- Etc.

e.g.

<https://docs.agno.com/tools/toolkits/toolkits>



Agent Scaffold

```
from agno.agent import Agent          # Core agent class
from agno.models.openai import OpenAIChat  # Wrapper around OpenAI chat models
from agno.tools.duckduckgo import DuckDuckGoTools
from textwrap import dedent

agent = Agent(
    model=OpenAIChat(id="gpt-4o"),      # 1 LLM brain
    tools=[DuckDuckGoTools()],          # 2 Tool belt
    description=dedent("""\
        You are Professor X-1000, a distinguished AI research scientist with expertise
        in analyzing and synthesizing complex information. Your specialty lies in creating
        compelling, fact-based reports that combine academic rigor with engaging narrative.
        Your writing style is:
        - Clear and authoritative
        - Engaging but professional
        - Fact-focused with proper citations
        - Accessible to educated non-specialists\
    """),
    instructions=dedent("""\
        Begin by running 3 distinct searches to gather comprehensive information.
        Analyze and cross-reference sources for accuracy and relevance.
        Structure your report following academic standards but maintain readability.
        Include only verifiable facts with proper citations.
        Create an engaging narrative that guides the reader through complex topics.
        End with actionable takeaways and future implications.\
    """),
    markdown=True,                      # Pretty output
    show_tool_calls=True,               # Expose what the agent does under the hood
    add_datetime_to_instructions=True,   # Embed today's date for temporal grounding
)

# ----- Example -----
if __name__ == "__main__":
    agent.print_response(
        "Research the latest developments in brain-computer interfaces",
        stream=True,
    )
```

System Prompt(s)

In Agno you pass
description (persona) +
instructions (strategy) as
two separate strings to
tweak them independently.



Agent Scaffold

```
from agno.agent import Agent          # Core agent class
from agno.models.openai import OpenAIChat  # Wrapper around OpenAI chat models
from agno.tools.duckduckgo import DuckDuckGoTools
from textwrap import dedent

agent = Agent(
    model=OpenAIChat(id="gpt-4o"),      # 1 LLM brain
    tools=[DuckDuckGoTools()],          # 2 Tool belt
    description=dedent("""\
        You are Professor X-1000, a distinguished AI research scientist with expertise
        in analyzing and synthesizing complex information. Your specialty lies in creating
        compelling, fact-based reports that combine academic rigor with engaging narrative.
        Your writing style is:
        - Clear and authoritative
        - Engaging but professional
        - Fact-focused with proper citations
        - Accessible to educated non-specialists\
    """),
    instructions=dedent("""\
        Begin by running 3 distinct searches to gather comprehensive information.
        Analyze and cross-reference sources for accuracy and relevance.
        Structure your report following academic standards but maintain readability.
        Include only verifiable facts with proper citations.
        Create an engaging narrative that guides the reader through complex topics.
        End with actionable takeaways and future implications.\
    """),
    markdown=True,                      # Pretty output
    show_tool_calls=True,               # Expose what the agent does under the hood
    add_datetime_to_instructions=True,   # Embed today's date for temporal grounding
)

# ----- Example -----
if __name__ == "__main__":
    agent.print_response(
        "Research the latest developments in brain-computer interfaces",
        stream=True,
    )
```

Additional Options

Many options exist



Agent Scaffold

```
from agno.agent import Agent          # Core agent class
from agno.models.openai import OpenAIChat  # Wrapper around OpenAI chat models
from agno.tools.duckduckgo import DuckDuckGoTools
from textwrap import dedent

agent = Agent(
    model=OpenAIChat(id="gpt-4o"),      # 1 LLM brain
    tools=[DuckDuckGoTools()],          # 2 Tool belt
    description=dedent("""\
        You are Professor X-1000, a distinguished AI research scientist with expertise
        in analyzing and synthesizing complex information. Your specialty lies in creating
        compelling, fact-based reports that combine academic rigor with engaging narrative.
        Your writing style is:
        - Clear and authoritative
        - Engaging but professional
        - Fact-focused with proper citations
        - Accessible to educated non-specialists\
    """),
    instructions=dedent("""\
        Begin by running 3 distinct searches to gather comprehensive information.
        Analyze and cross-reference sources for accuracy and relevance.
        Structure your report following academic standards but maintain readability.
        Include only verifiable facts with proper citations.
        Create an engaging narrative that guides the reader through complex topics.
        End with actionable takeaways and future implications.\
    """),
    markdown=True,                      # Pretty output
    show_tool_calls=True,               # Expose what the agent does under the hood
    add_datetime_to_instructions=True,   # Embed today's date for temporal grounding
)

# ----- Example -----
if __name__ == "__main__":
    agent.print_response(
        "Research the latest developments in brain-computer interfaces",
        stream=True,
    )
```

} **User Prompt**

Question that needs
answering

Agent Final Tips

Context windows

Every model has a token limit (~128k for GPT-4o).

Keep it small!

- Short system prompt (≤ 50 tokens).
- Use bullet lists instead of paragraphs.
- Clip extra history you don't need.

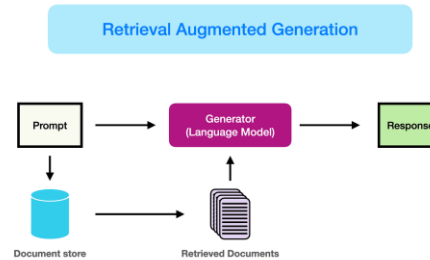
Breakout A

- 15 minutes
- **Task:** Securities Forecasting -
 - Ensure agent runs successfully
 - Add additional tooling, edit prompts
 - Forecast a stocks performance
 - Discuss results
- **Goal:** Tooling and Prompting Config
 - Evaluate different prompting techniques
 - Evaluate how tooling changes results
- **Feedback:** How can we evaluate the performance of the responses?



RAGs to Riches

RAG (Retrieval-Augmented Generation)



Motto: *“If the knowledge isn’t in the model, fetch it.”*

Agent Scaffold

```
from agno.agent import Agent
from agno.models.openai import OpenAIChat
from agno.knowledge.pdf import PDFKnowledgeBase
from agno.vectordb.pgvector import PgVector, SearchType
```

} Imports

```
DB_URL = "postgresql+psycpg://ai:ai@localhost:5532/ai"
```

```
knowledge_base = PDFUrlKnowledgeBase(
    urls=["$RETRIEVED FROM ABOVE"],
    vector_db=PgVector(table_name="financial_reports", db_url=DB_URL, search_type=SearchType.hybrid),
)
knowledge_base.load(load=True, upsert=True)
```

```
agent = Agent(
    model=OpenAIChat(id="gpt-4o"),
    knowledge=knowledge_base,
    search_knowledge=True,
    markdown=True,
)
```

```
agent.print_response("What what NVDA's quarterly revenue in 2024?", stream=True)
```

Agent Scaffold

```
from agno.agent import Agent
from agno.models.openai import OpenAIChat
from agno.knowledge.pdf import PDFKnowledgeBase
from agno.vectordb.pgvector import PgVector, SearchType

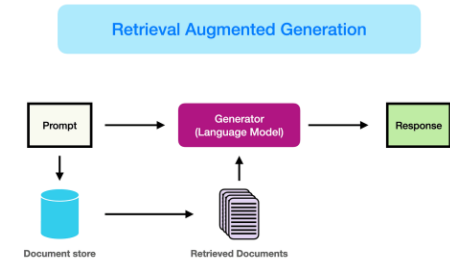
DB_URL = "postgresql+psycpg://ai:ai@localhost:5532/ai"

knowledge_base = PDFUrlKnowledgeBase(
    urls=["$RETRIEVED FROM ABOVE"],
    vector_db=PgVector(table_name="financial_reports", db_url=DB_URL, search_type=SearchType.hybrid),
)
knowledge_base.load(load=True, upsert=True)

agent = Agent(
    model=OpenAIChat(id="gpt-4o"),
    knowledge=knowledge_base,
    search_knowledge=True,
    markdown=True,
)

agent.print_response("What what NVDA's quarterly revenue in 2024?", stream=True)
```

} Database



Agent Scaffold

```
from agno.agent import Agent
from agno.models.openai import OpenAIChat
from agno.knowledge.pdf import PDFKnowledgeBase
from agno.vectordb.pgvector import PgVector, SearchType

DB_URL = "postgresql+psycpg://ai:ai@localhost:5532/ai"

knowledge_base = PDFUrlKnowledgeBase(
    urls=["$RETRIEVED FROM ABOVE"],
    vector_db=PgVector(table_name="financial_reports", db_url=DB_URL, search_type=SearchType.hybrid),
)
knowledge_base.load(load=True, upsert=True)

agent = Agent(
    model=OpenAIChat(id="gpt-4o"),
    knowledge=knowledge_base,
    search_knowledge=True,
    markdown=True,
)

agent.print_response("What what NVDA's quarterly revenue in 2024?", stream=True)
```

Knowledge Base

PDFUrlKnowledgeBase:
PDF information is
tokenized and stored in
database

Other KnowledgeBase
plugins exist

Agent Scaffold

```
from agno.agent import Agent
from agno.models.openai import OpenAIChat
from agno.knowledge.pdf import PDFKnowledgeBase
from agno.vectordb.pgvector import PgVector, SearchType

DB_URL = "postgresql+psycopg://ai:ai@localhost:5532/ai"

knowledge_base = PDFUrlKnowledgeBase(
    urls=["$RETRIEVED FROM ABOVE"],
    vector_db=PgVector(table_name="financial_reports", db_url=DB_URL, search_type=SearchType.hybrid),
)
knowledge_base.load(load=True, upsert=True)

agent = Agent(
    model=OpenAIChat(id="gpt-4o"),
    knowledge=knowledge_base,
    search_knowledge=True,
    markdown=True,
)

agent.print_response("What what NVDA's quarterly revenue in 2024?", stream=True)
```

} **Agent Specification**

Breakout B

- 15 minutes
- **Task:** Securities Forecasting -
 - Ensure agent runs successfully
 - Use prompt engineering &
 - Integrate recent quarterly reports to improve forecast
 - Discuss results
- **Goal:** RAG Configuration
 - Evaluate how retrieved information informs results
- **Feedback:** What other kinds of information may help to inform the Agent?



Closing & Next Steps

Closing:

- Thank you!
- Objectives summary
- Q/A
 - AI, Software / Finance Industry, Our Work, etc.
 - We're (mostly) an open book! 😊

Objectives

- Gain familiarity with an Agent Framework
- Understand different prompt techniques
- Enhance responses using Retrieval Augmented Generation