Large Language Models in finance

Laurent Sorber

Financieel Forum Vlaams-Brabant
Leuven, 20 November 2023

InvestSuite

Radix

Outline

- Who are InvestSuite & Radix
- What are Large Language Models
- The magic of LLMs
- Dispelling the magic of LLMs
- Applications of LLMs
- What does the future hold

Copyright © RADIX.AI BVBA. All rights reserved.

Outline

- **Who are InvestSuite & Radix**
- What are Large Language Models
- The magic of LLMs
- Dispelling the magic of LLMs
- Applications of LLMs
- What does the future hold

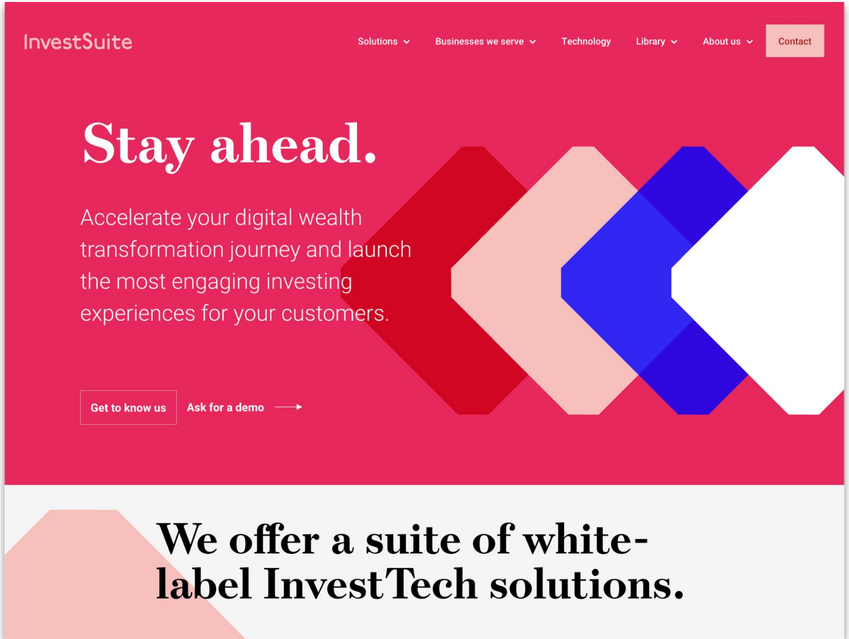
Radix — Copyright © RADIX.AI BVBA. All rights reserved.

InvestSuite

Radix

A suite of investtech products with which we want to make investing more accessible to a wider audience

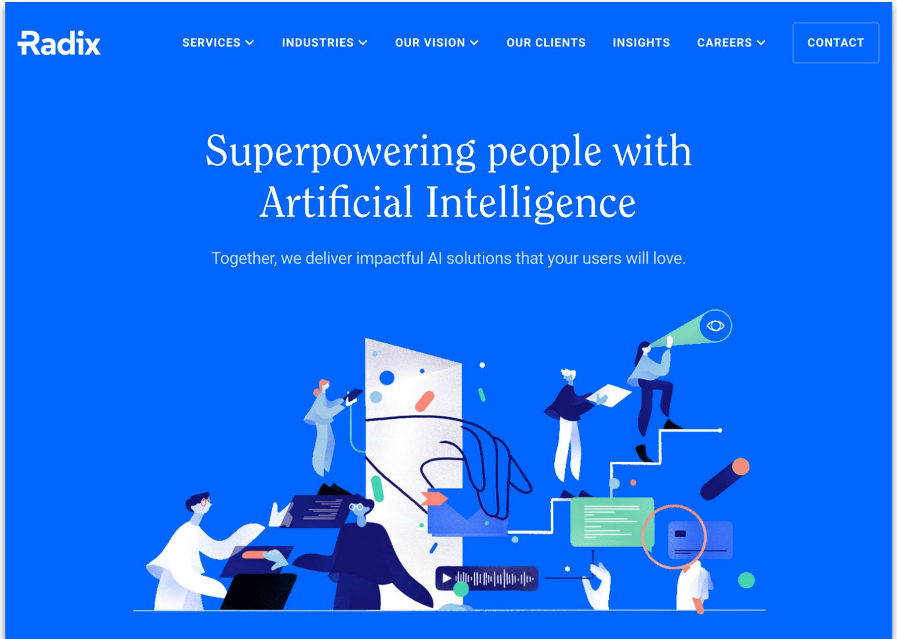
Bespoke AI solutions with which we want to superpower people for clients across industries and AI domains



Stay ahead.

Accelerate your digital wealth transformation journey and launch the most engaging investing experiences for your customers.

We offer a suite of white-label InvestTech solutions.



Superpowering people with Artificial Intelligence

Together, we deliver impactful AI solutions that your users will love.

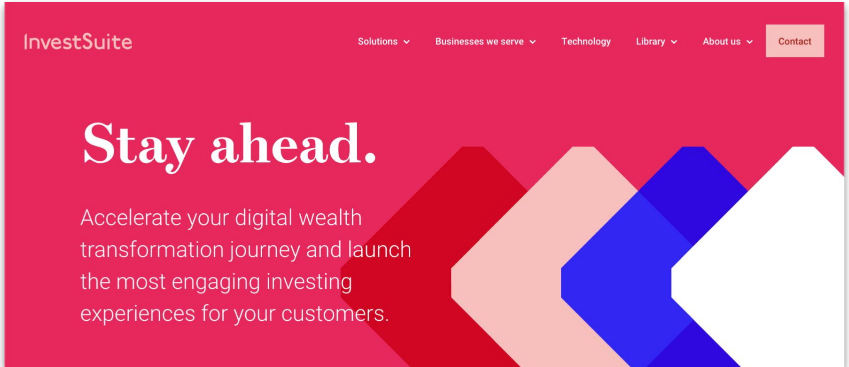
A suite of investtech products with which we want to make investing more accessible to a wider audience

Founded 2018

Offices in Leuven, CH, UK

50+ colleagues

Sales presence in BE, DE, NL, USA



Stay ahead.

Accelerate your digital wealth transformation journey and launch the most engaging investing experiences for your customers.

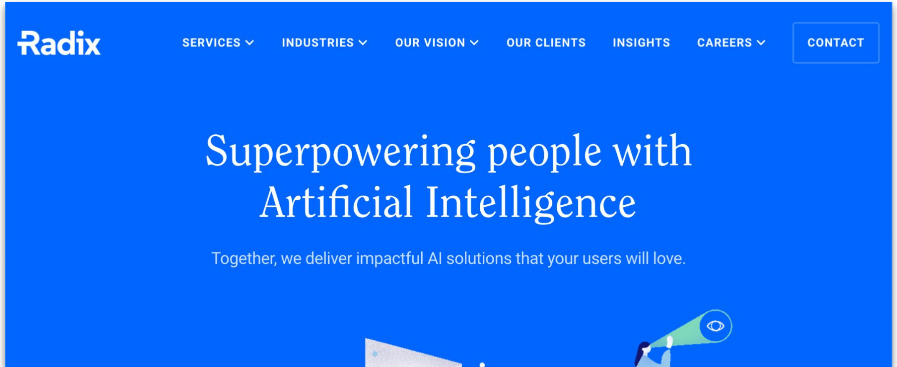
Bespoke AI solutions with which we want to superpower people for clients across industries and AI domains

Founded 2018

Offices in Brussels & Ghent

50+ colleagues

Active in BE, NL, DE, IT, Nordics, USA



Superpowering people with Artificial Intelligence

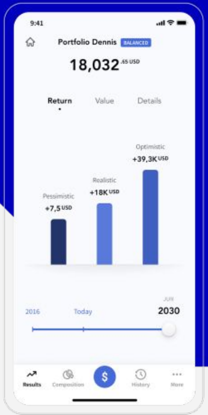
Together, we deliver impactful AI solutions that your users will love.

InvestSuite

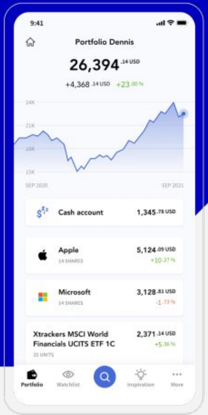
Investment Platforms

Portfolio Construction

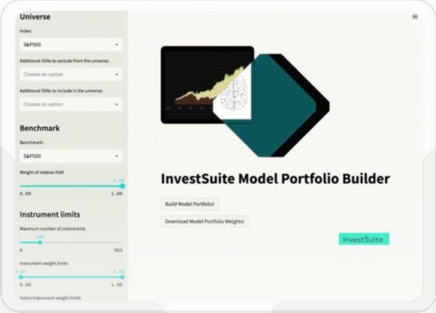
Investment Insights & Reporting



Robo Advisor



Self Investor



Universe

Index: S&P500

Additional ISINs to exclude from the universe: Choose an option

Additional ISINs to include in the universe: Choose an option

Benchmark

Benchmark: S&P500

Weight of relative VaR

Instrument limits

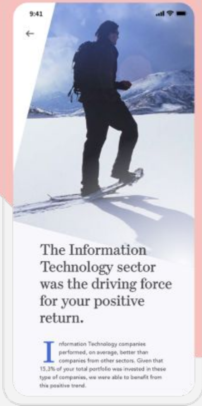
Maximum number of instruments

Instrument weight limits

Active instrument weight limits

InvestSuite Model Portfolio Builder

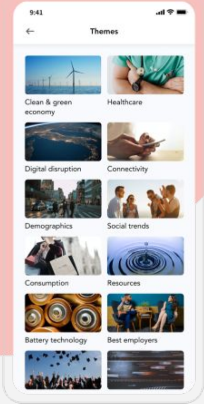
Portfolio Optimizer



The Information Technology sector was the driving force for your positive return.

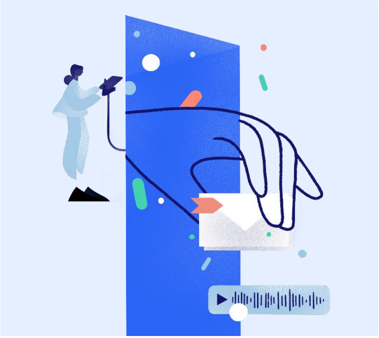
Information Technology companies performed, on average, better than companies from other sectors. Given that 15.3% of your total portfolio was invested in these type of companies, we were able to benefit from this positive trend.

StoryTeller



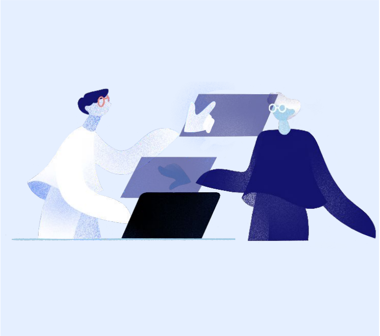
Insight API

Radix



AI solutions

- AI modelling
- Software engineering
- Cloud infrastructure
- MLOps
- Data annotation



Knowledge transfer

- Training
- Coaching
- Team extensions

AI strategy

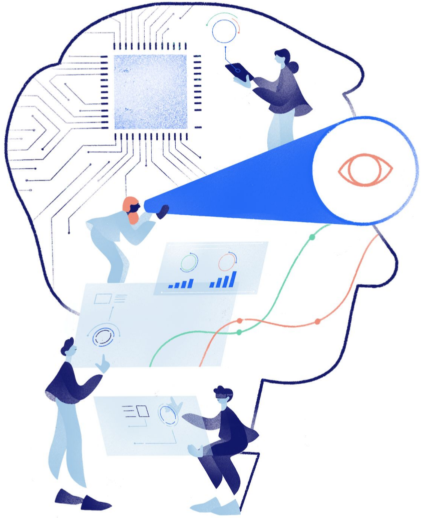
- Discover the best opportunities in AI
- Value & effort estimation of AI opportunities

Research

- Develop new AI models
- Explore new applications
- Validate research results

Radix — Copyright © RADIX.AI BVBA. All rights reserved.

Radix



Natural Language Processing 🧠

- Text generation
- Question answering
- Summarization
- Named entity recognition
- Classification
- Duplicate detection
- Information retrieval

Tabular Data 📊

- Classification
- Regression
- Anomaly detection
- Explainable prediction
- Confidence intervals
- Matrix imputation
- Feature engineering
- Data visualization

Computer Vision 👁

- Image detection
- Object detection
- Semantic segmentation
- Instance segmentation
- Real-time detection
- Hyperspectral imaging
- Super-resolution
- Optical Character Recognition

Planning 📆

- Scheduling optimization
- Assignment optimization
- Inventory optimization
- Routing optimization

Time Series 📈

- Forecasting
- Anomaly detection
- Classification
- Confidence intervals

Speech 🎤

- Transcription
- Speech recognition
- Voice cloning
- Speaker identification

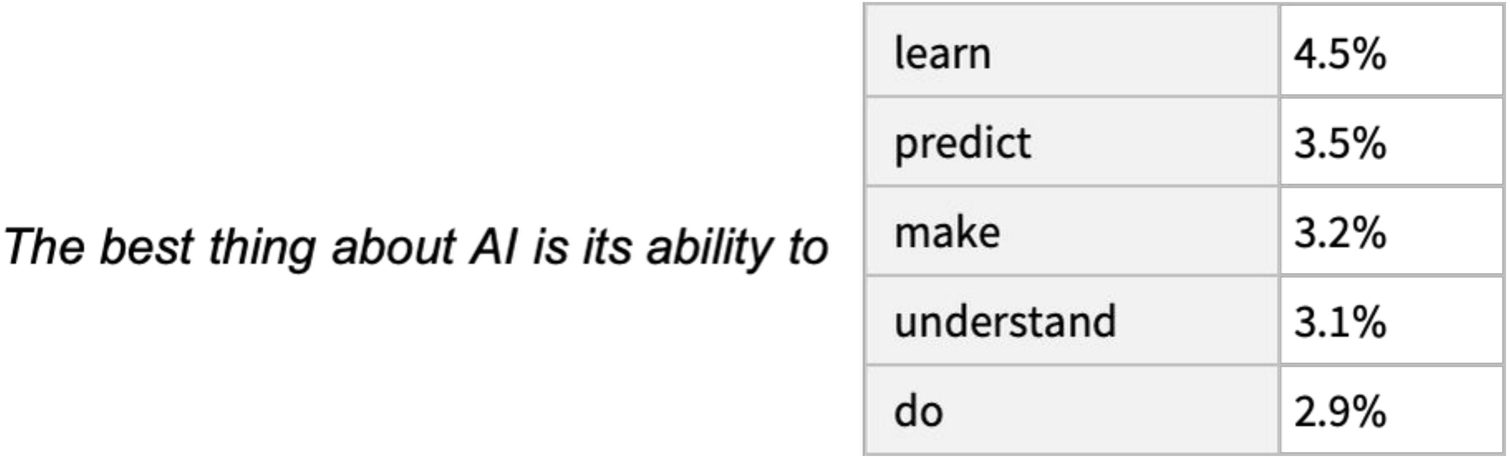
Radix — Copyright © RADIX.AI BVBA. All rights reserved.

Outline

- ● Who are InvestSuite & Radix
- ● **What are Large Language Models**
- ● The magic of LLMs
- ● Dispelling the magic of LLMs
- ● Applications of LLMs
- ● What does the future hold

Copyright © RADIX.AI BVBA. All rights reserved.

What is an LLM?



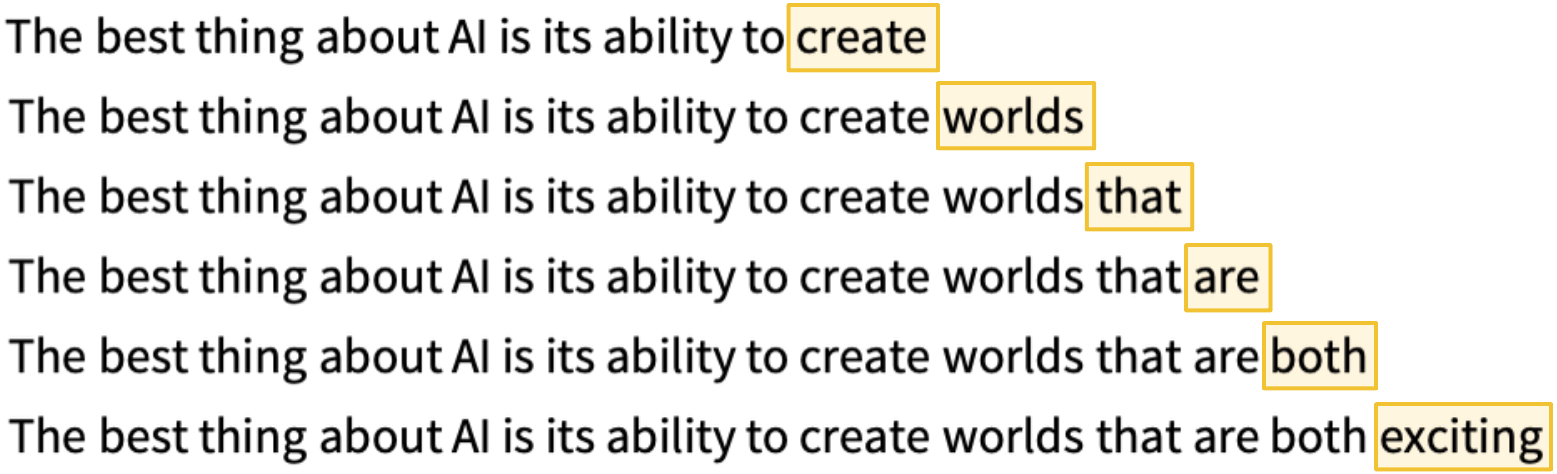
The best thing about AI is its ability to

learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

[1] Image source: https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/

Copyright © RADIX.AI BVBA. All rights reserved.

What is an LLM?



The best thing about AI is its ability to create

The best thing about AI is its ability to create worlds

The best thing about AI is its ability to create worlds that

The best thing about AI is its ability to create worlds that are

The best thing about AI is its ability to create worlds that are both

The best thing about AI is its ability to create worlds that are both exciting

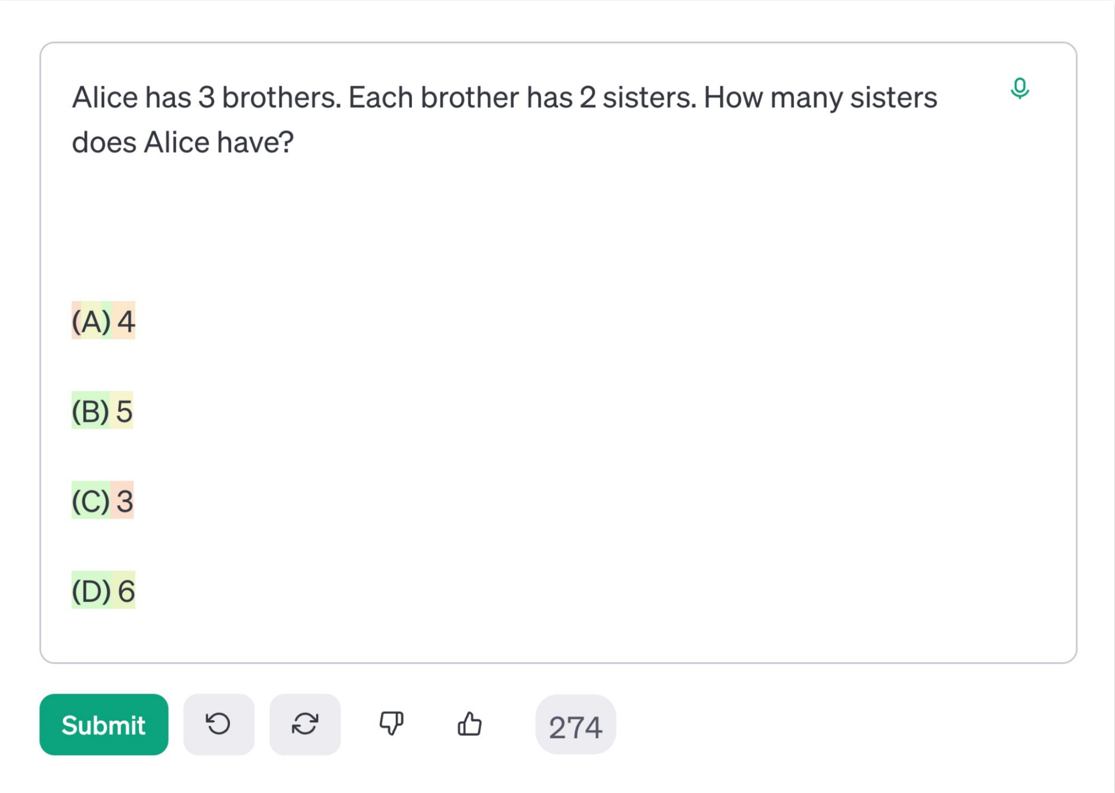
---

[1] Image source: https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/

Copyright © RADIX.AI BVBA. All rights reserved.

Base LLM:

Completes documents



Alice has 3 brothers. Each brother has 2 sisters. How many sisters does Alice have?

(A) 4

(B) 5

(C) 3

(D) 6

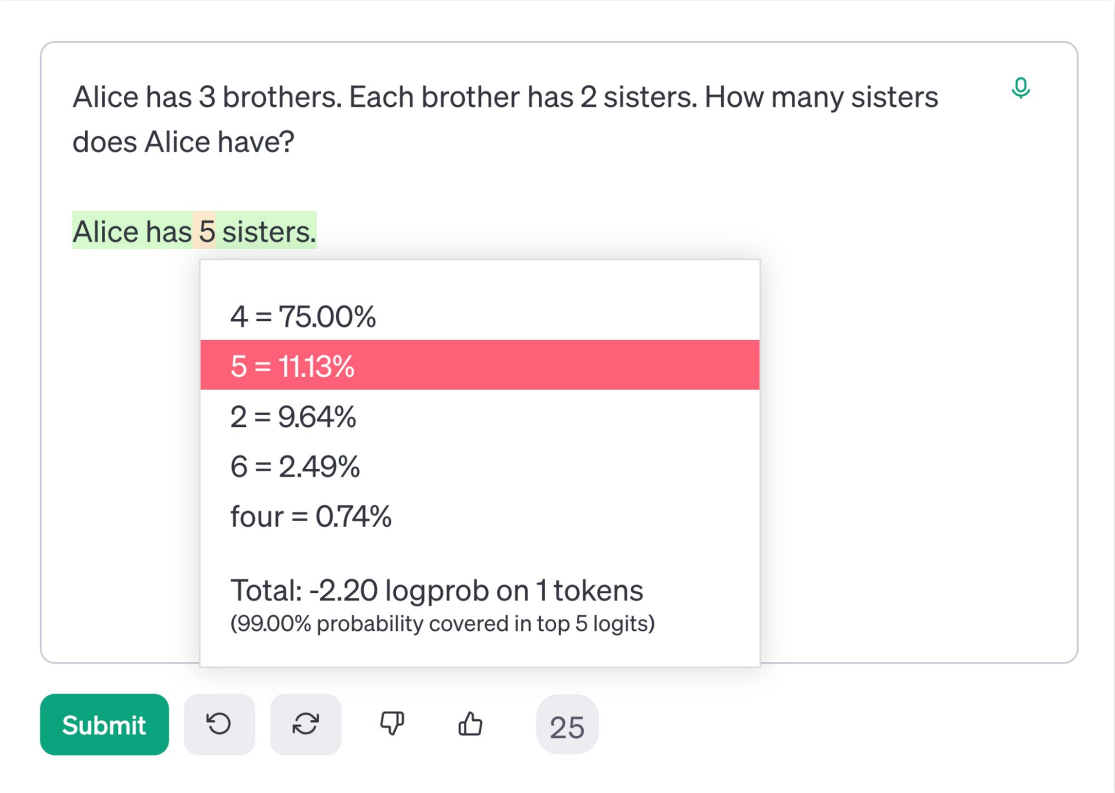
---

[1] OpenAI Playground with base GPT-3: https://platform.openai.com/playground?mode=complete&model=davinci

Radix – Copyright © RADIX.AI BVBA. All rights reserved.

Instruction tuned LLM:

Answers questions



Alice has 3 brothers. Each brother has 2 sisters. How many sisters does Alice have?

Alice has 5 sisters.

4 = 75.00%
5 = 11.13%
2 = 9.64%
6 = 2.49%
four = 0.74%

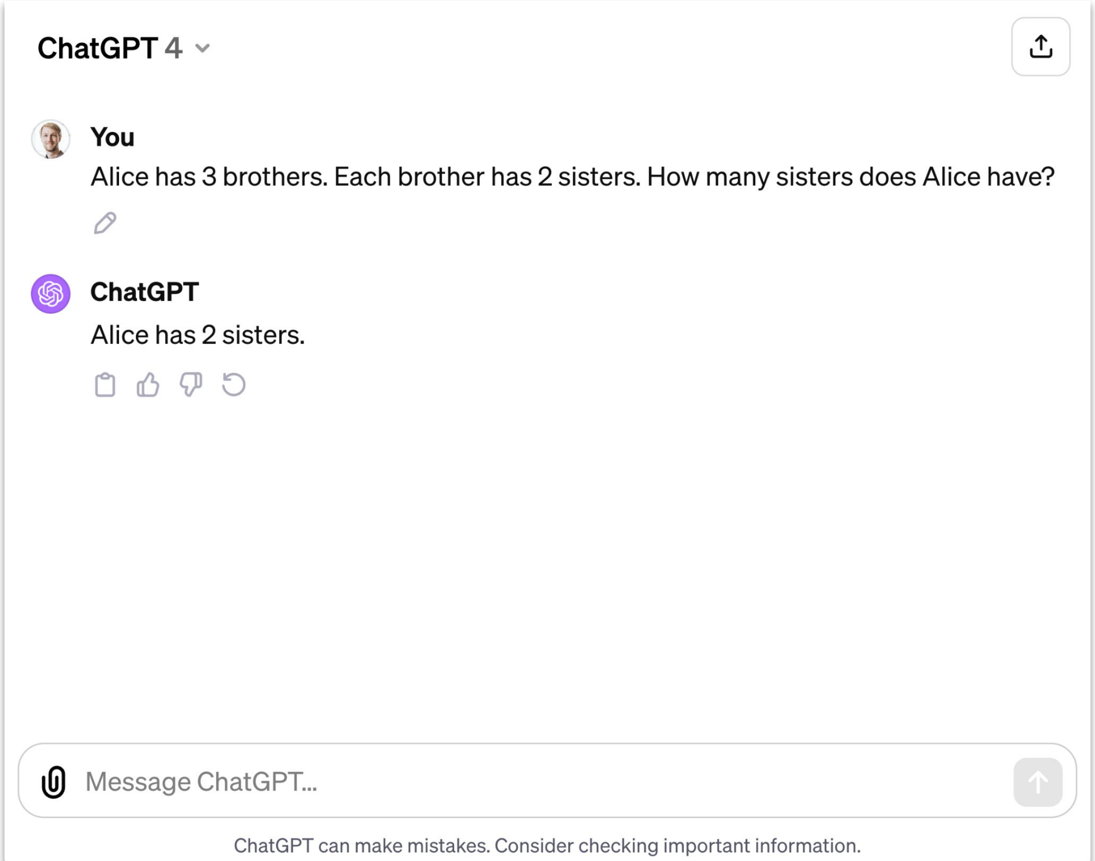
Total: -2.20 logprob on 1 tokens
(99.00% probability covered in top 5 logits)

[1] OpenAI Playground with instruction tuned GPT-3.5: https://platform.openai.com/playground?mode=complete&model=text-davinci-003

Copyright © RADIX.AI BVBA. All rights reserved.

ChatGPT:

A web app on top of GPT-4



You

Alice has 3 brothers. Each brother has 2 sisters. How many sisters does Alice have?

ChatGPT

Alice has 2 sisters.

ChatGPT can make mistakes. Consider checking important information.

---

[1] ChatGPT: https://chat.openai.com

Radix — Copyright © RADIX.AI BVBA. All rights reserved.

Outline

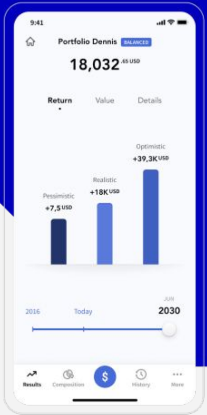
- Who are InvestSuite & Radix
- What are Large Language Models
- **The magic of LLMs**
- Dispelling the magic of LLMs
- Applications of LLMs
- What does the future hold

Radix — Copyright © RADIX.AI BVBA. All rights reserved.

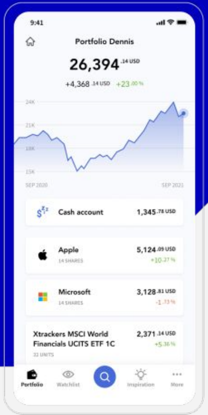
Demo

InvestSuite

Investment Platforms



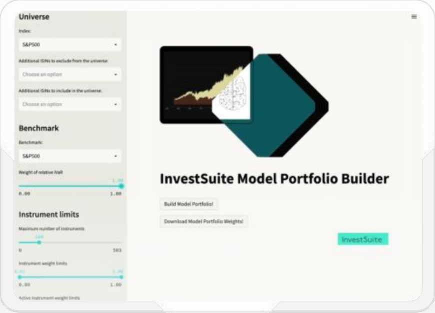
Robo Advisor



Cash account	1,345 .78 USD
Apple 14 SHARES	5,124 .09 USD +10.27 %
Microsoft 14 SHARES	3,128 .81 USD -1.73 %
Xtrackers MSCI World Financials UCITS ETF 1C 32 UNITS	2,371 .14 USD +5.36 %

Self Investor

Portfolio Construction



Universe

Index: S&P500

Additional ISINs to exclude from the universe: Choose an option

Additional ISINs to include in the universe: Choose an option

Benchmark

Benchmark: S&P500

Weight of relative WaR

Instrument limits

Maximum number of instruments

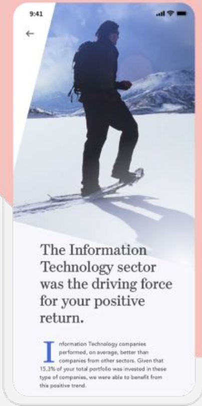
Instrument weight limits

Active instrument weight limits

0.00
1.00
0
503
0.00
1.00
InvestSuite Model Portfolio Builder
Build Model Portfolio!
Download Model Portfolio Weights!
InvestSuite

Portfolio Optimizer

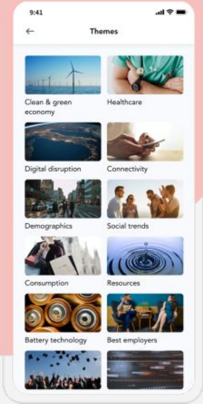
Investment Insights & Reporting



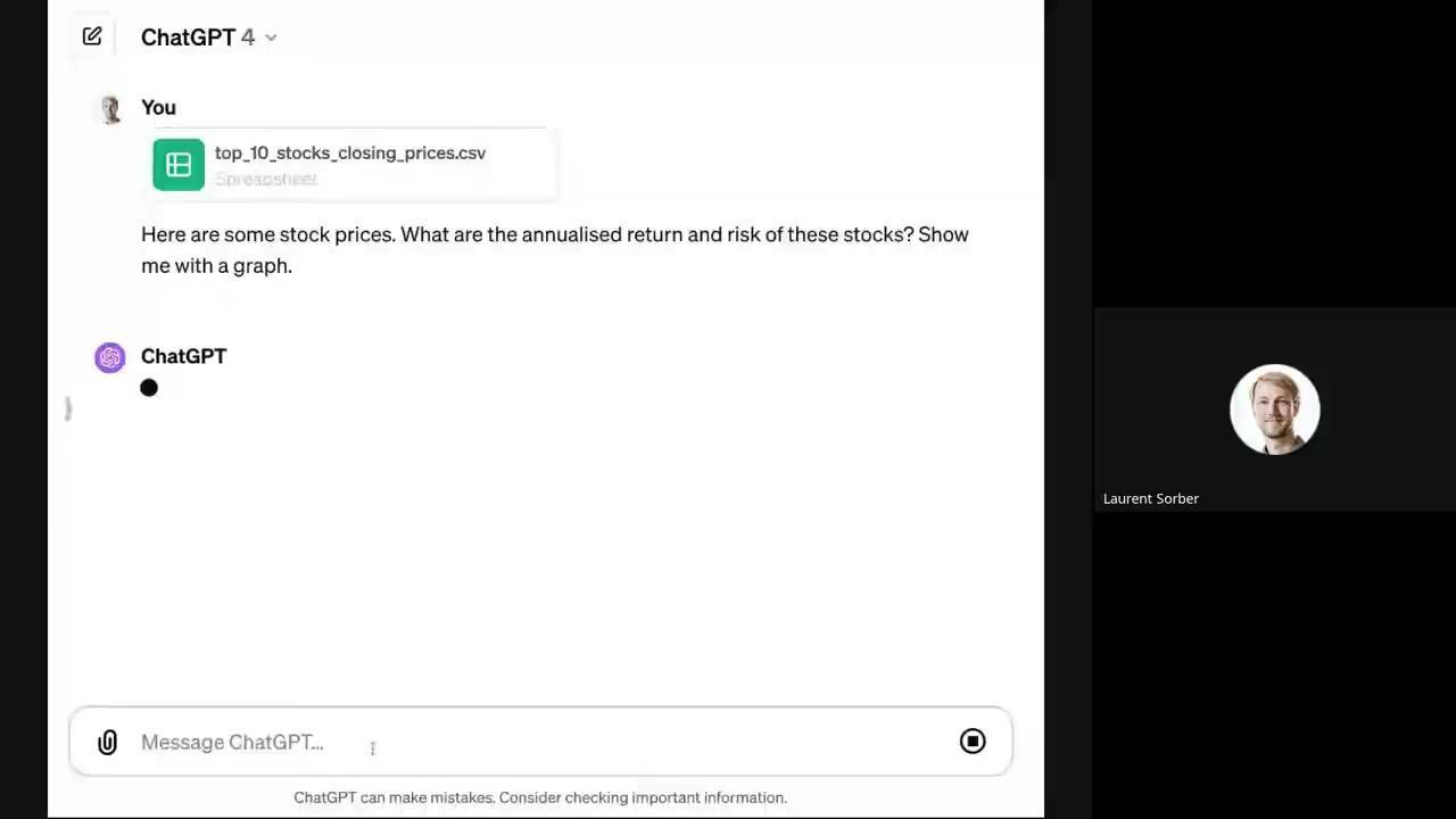
The Information Technology sector was the driving force for your positive return.

Information Technology companies performed, on average, better than companies from other sectors. Given that 15,3% of your total portfolio was invested in these type of companies, we were able to benefit from this positive trend.

StoryTeller



Insight API



Here are some stock prices. What are the annualised return and risk of these stocks? Show me with a graph.

ChatGPT can make mistakes. Consider checking important information.

The keys to success

Instruction tuning

Scaling to 1T tokens (≈ words)

Outline

- Who are InvestSuite & Radix
- What are Large Language Models
- The magic of LLMs
- **Dispelling the magic of LLMs**
- Applications of LLMs
- What does the future hold

All LLMs are Transformers

arXiv:1706.03762v7 [cs.CL] 2 Aug 2023

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

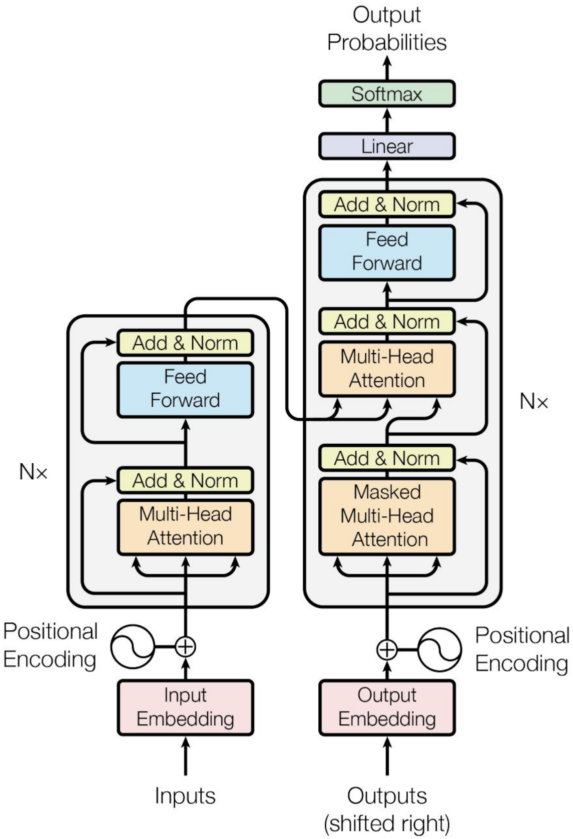
---

[*] Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

[†] Work performed while at Google Brain.

[‡] Work performed while at Google Research.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.



Radix — Copyright © RADIX.AI BVBA. All rights reserved.

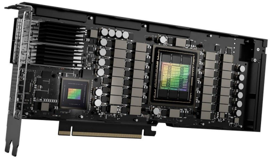
Input text

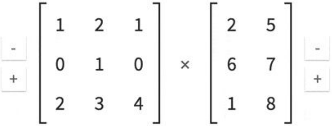
Weights & Biases is the

Tokenize

[1135, 2337, 1222, 8436, 1386, 318, 262]

LLM





$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \\ 2 & 3 & 4 \end{bmatrix} \times \begin{bmatrix} 2 & 5 \\ 6 & 7 \\ 1 & 8 \end{bmatrix}$$

Output probabilities

...	...
We	0.001
ights	0.001
&	0.001
the	0.001
machine	0.15
learning	0.001
platform	0.2

Sample

machine

---

[1] LLM image source: https://wandb.ai/darek/llmapps/reports/A-Gentle-Introduction-to-LLM-APIs--Vmlldzo0NjM0MTMz

[2] GPU image source: https://www.nvidia.com/en-us/data-center/h100/

[3] Matrix multiplication source: https://math.stackexchange.com/questions/3715598/when-should-hadamard-matrix-multiplication-be-used-vs-other-methods

What do LLMs imitate?

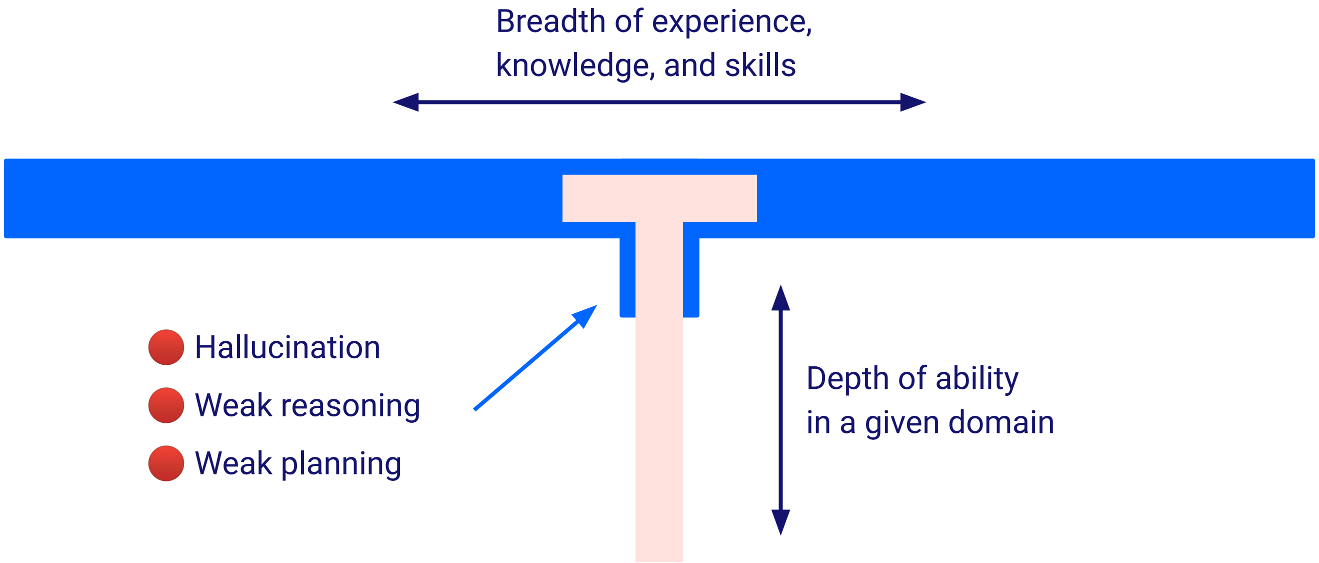
✅ Cognitive ability

❌ Consciousness

❌ Sentience

❌ Will

The T-shape of an LLM



Breadth of experience, knowledge, and skills

- 🔴 Hallucination
- 🔴 Weak reasoning
- 🔴 Weak planning

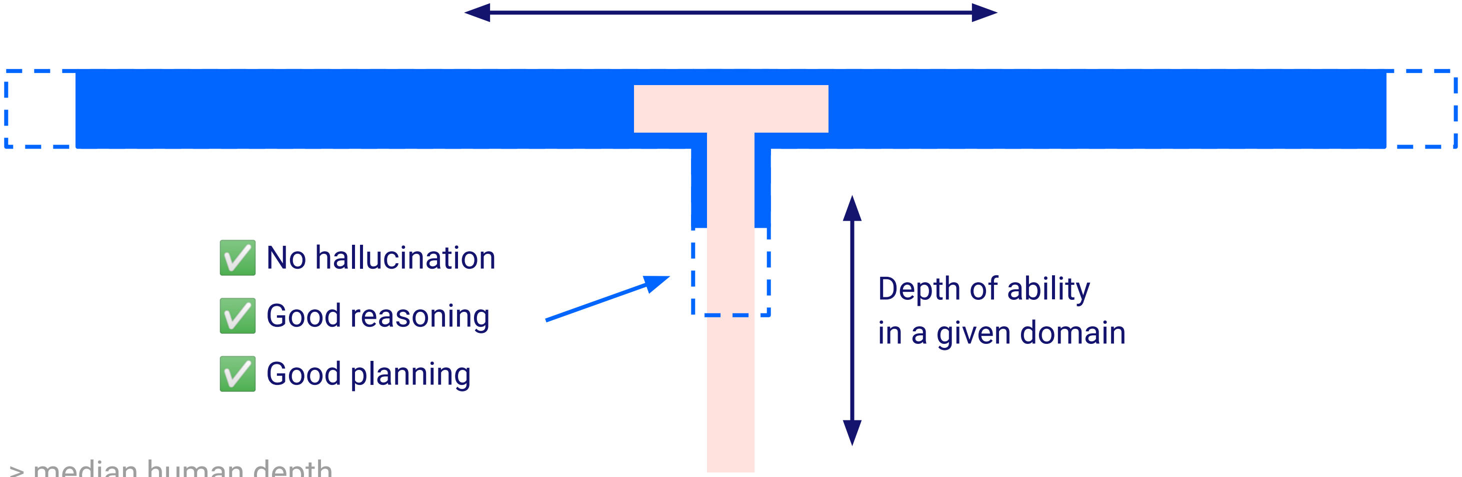
Depth of ability in a given domain

🟦 **LLM**

◻ **Human expert**

Radix — Copyright © RADIX.AI BVBA. All rights reserved.

The T-shape of an LLM



Breadth of experience,
knowledge, and skills

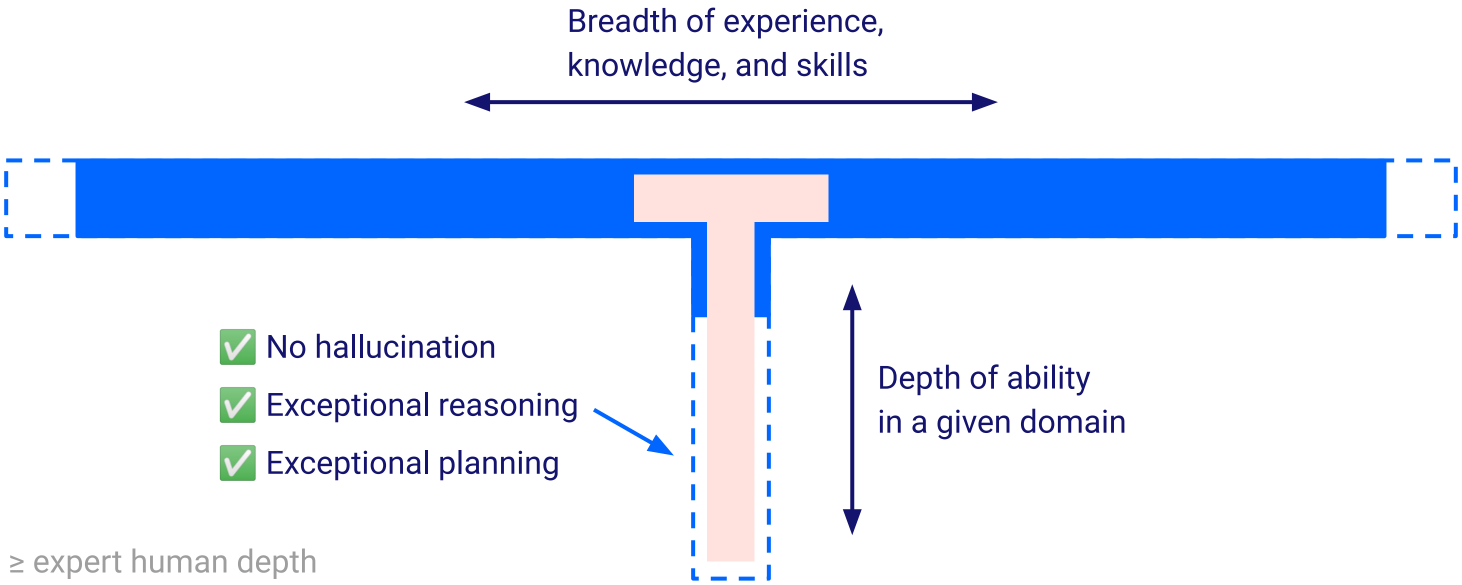
✅ No hallucination
✅ Good reasoning
✅ Good planning

Depth of ability
in a given domain

⬚ **AGI** ≥ median human depth
■ **LLM**
□ **Human expert**

Radix — Copyright © RADIX.AI BVBA. All rights reserved.

The T-shape of an LLM



Breadth of experience, knowledge, and skills

Depth of ability in a given domain

- ✅ No hallucination
- ✅ Exceptional reasoning
- ✅ Exceptional planning

⬚ **ASI** ≥ expert human depth

🟦 **LLM**

▨ **Human expert**

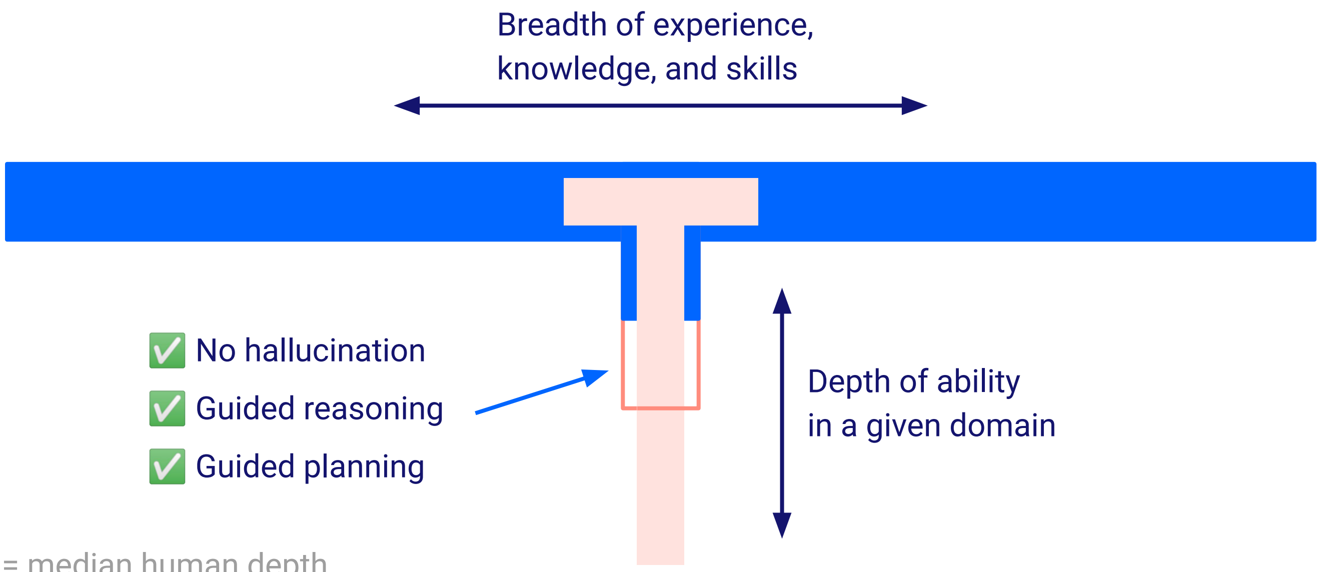
Radix — Copyright © RADIX.AI BVBA. All rights reserved.

Outline

- Who are InvestSuite & Radix
- What are Large Language Models
- The magic of LLMs
- Dispelling the magic of LLMs
- **Applications of LLMs**
- What does the future hold

Radix — Copyright © RADIX.AI BVBA. All rights reserved.

The T-shape of an LLM app



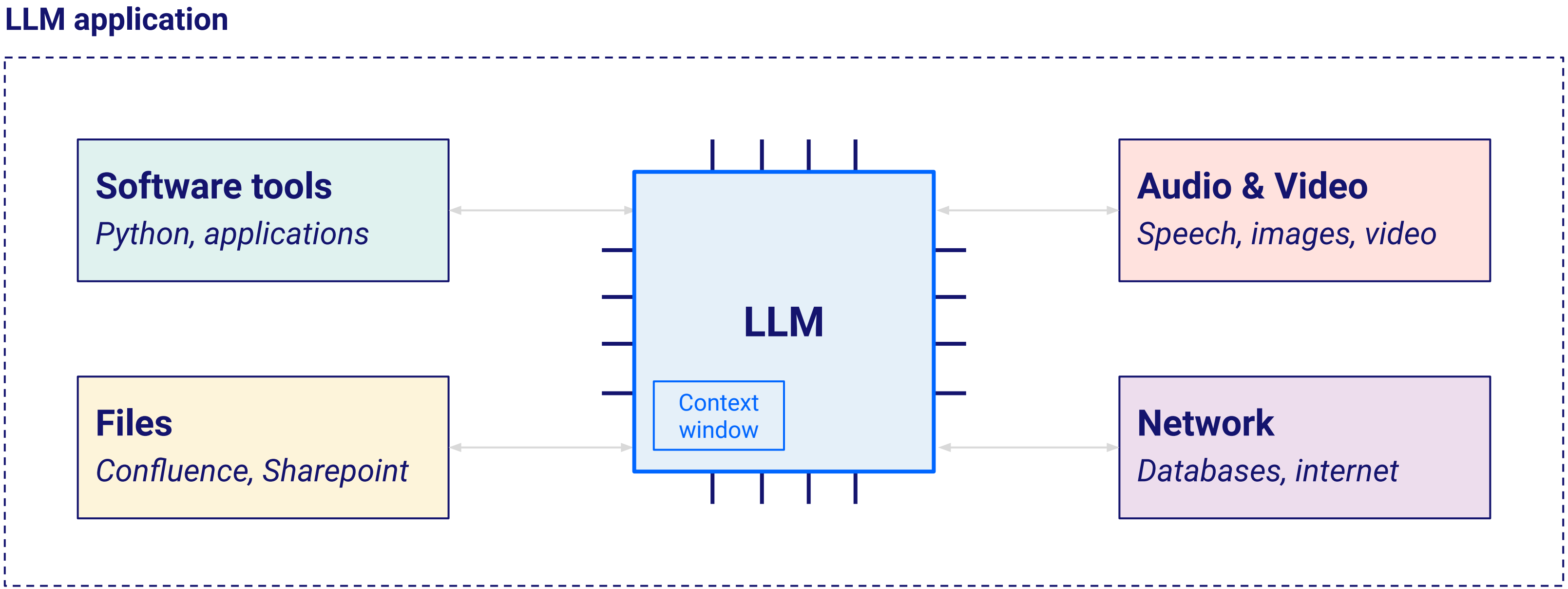
Breadth of experience, knowledge, and skills

✅ No hallucination
✅ Guided reasoning
✅ Guided planning

Depth of ability in a given domain

☐ **LLM app** = median human depth
🟦 **LLM**
▨ **Human expert**

Anatomy of an LLM app



LLM application

Software tools

Python, applications

Files

Confluence, Sharepoint

LLM

Context window

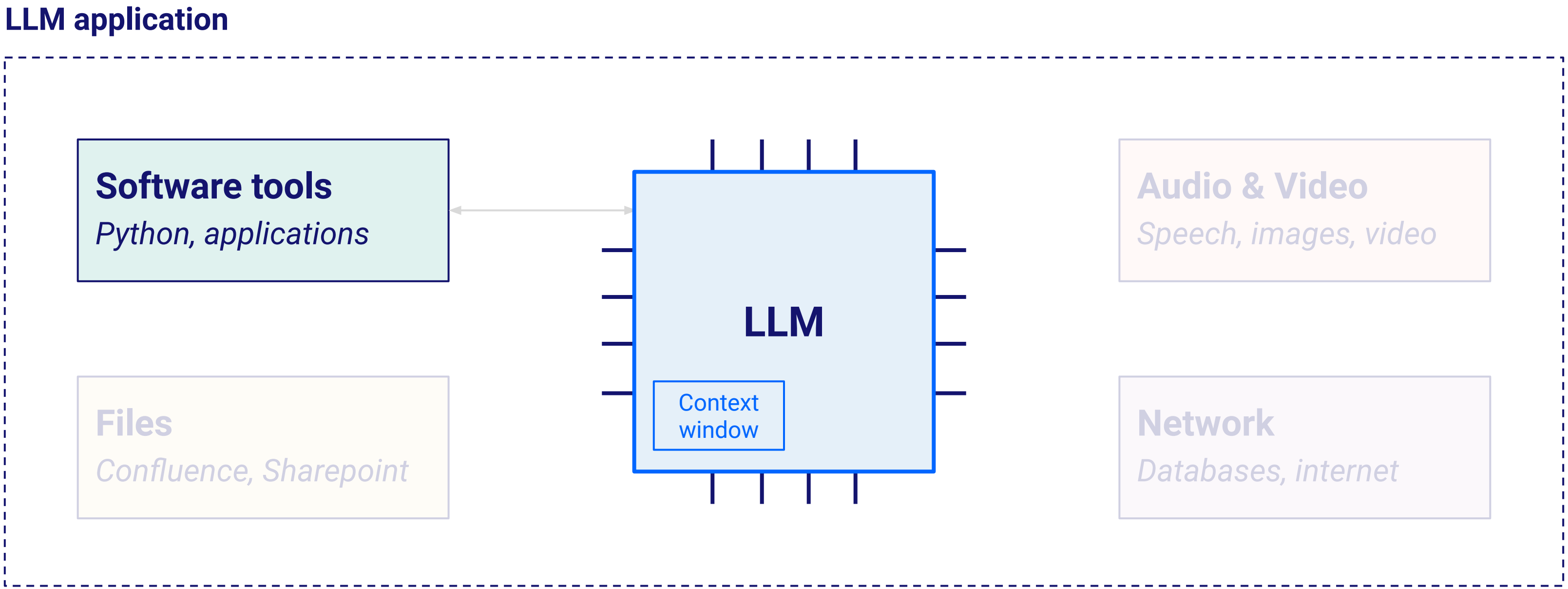
Audio & Video

Speech, images, video

Network

Databases, internet

LLM app: Portfolio Assistant



LLM application

Software tools

Python, applications

LLM

Context window

Audio & Video

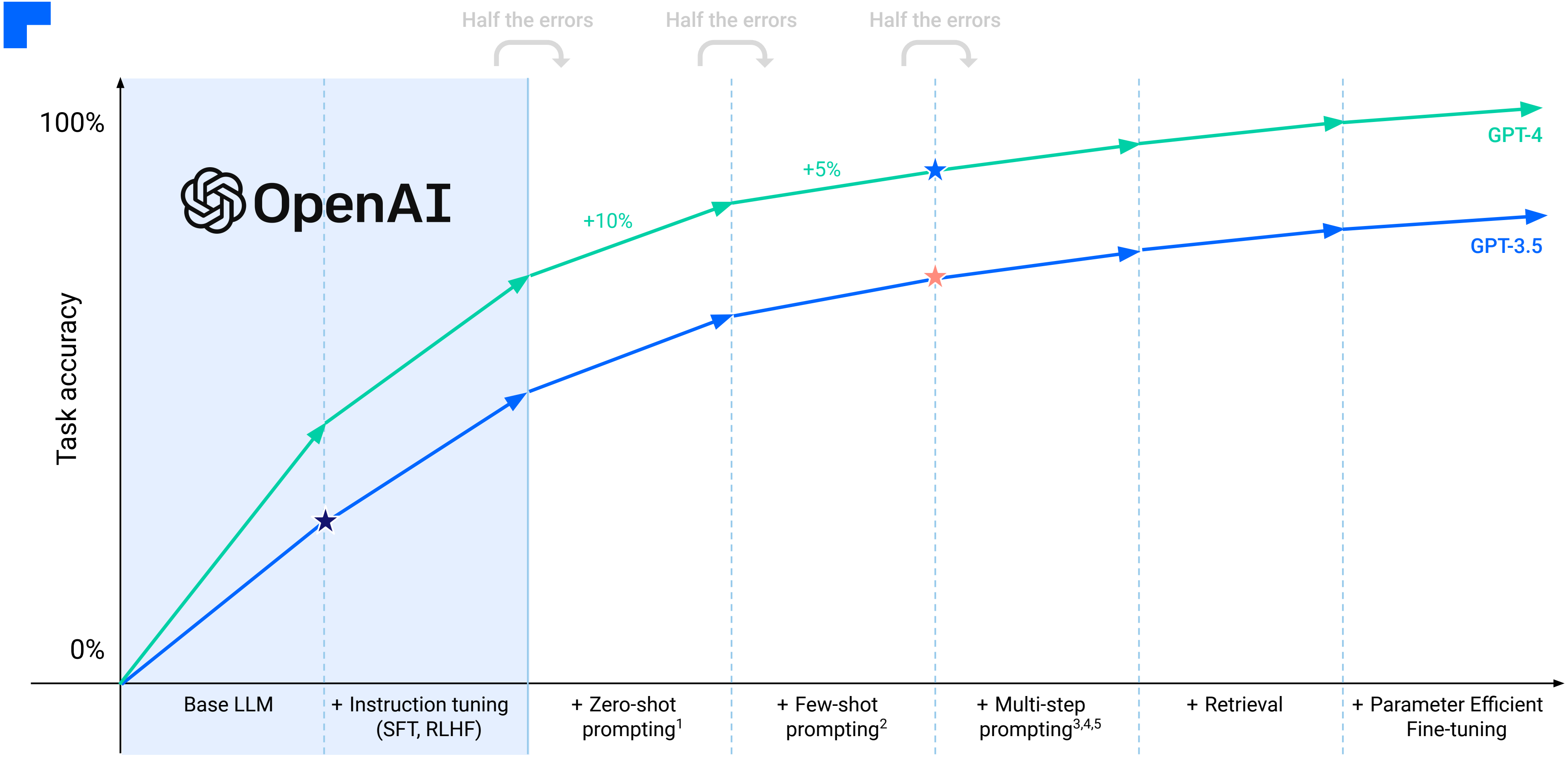
Speech, images, video

Files

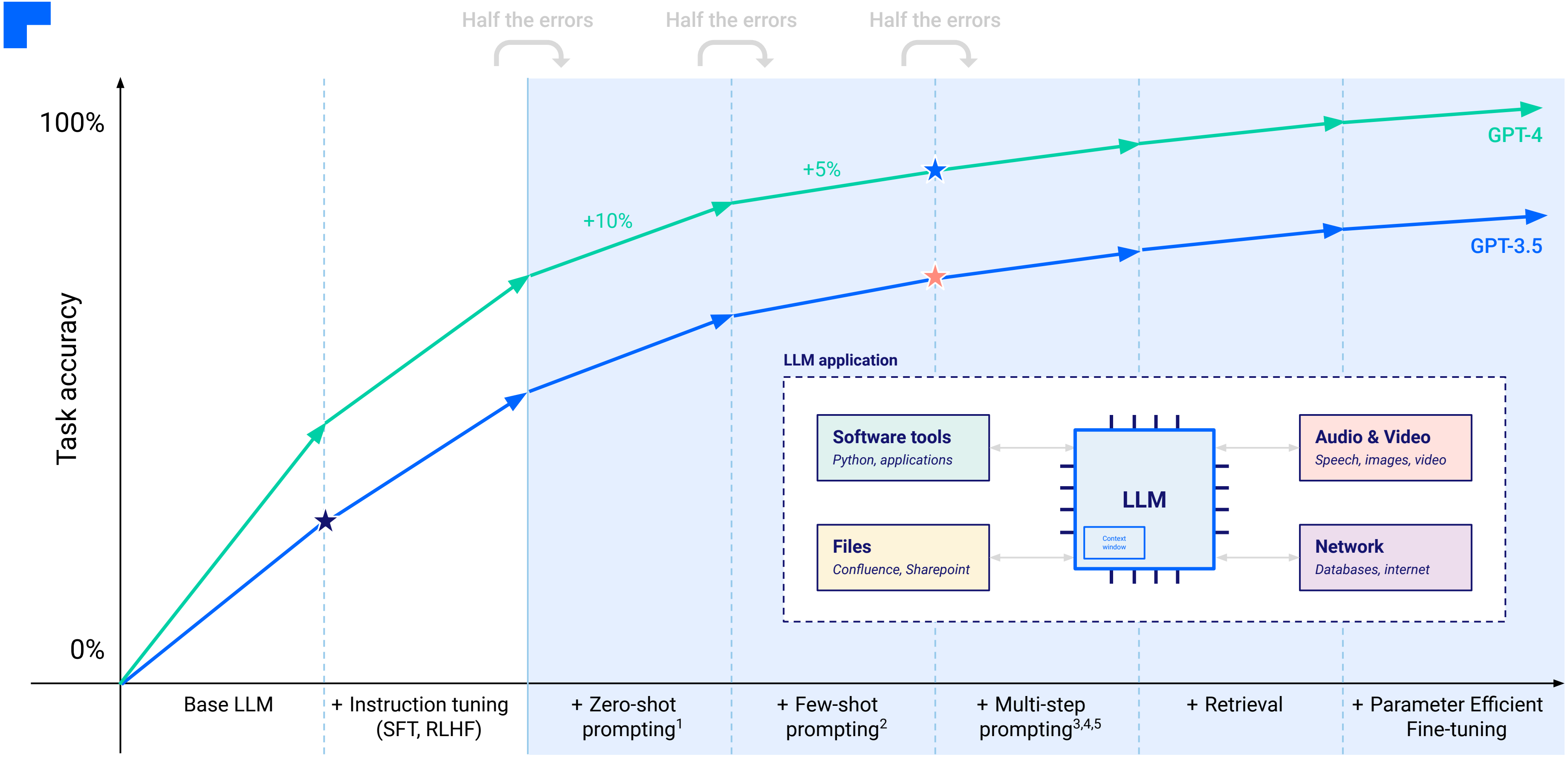
Confluence, Sharepoint

Network

Databases, internet



★ GPT-4 achieves a few-shot chain-of-thought performance on MMLU of 86%: https://arxiv.org/abs/2305.17306
★ GPT-3.5 achieves a few-shot chain-of-thought performance on MMLU of 67%: https://arxiv.org/abs/2305.17306
★ GPT-3 achieves a few-shot performance on MMLU of 44%: https://arxiv.org/pdf/2009.03300.pdf
[1] An automatically discovered chain-of-thought prompt generalizes to novel models and datasets: https://arxiv.org/pdf/2305.02897.pdf
[2] Language Models are Few-Shot Learners: https://arxiv.org/pdf/2005.14165.pdf
[3,4,5] Self-consistency: https://arxiv.org/pdf/2203.11171.pdf, Self-reflection: https://evjang.com/2023/03/26/self-reflection.html, Tree of Thoughts: https://arxiv.org/pdf/2305.10601.pdf



★ GPT-4 achieves a few-shot chain-of-thought performance on MMLU of 86%: https://arxiv.org/abs/2305.17306

★ GPT-3.5 achieves a few-shot chain-of-thought performance on MMLU of 67%: https://arxiv.org/abs/2305.17306

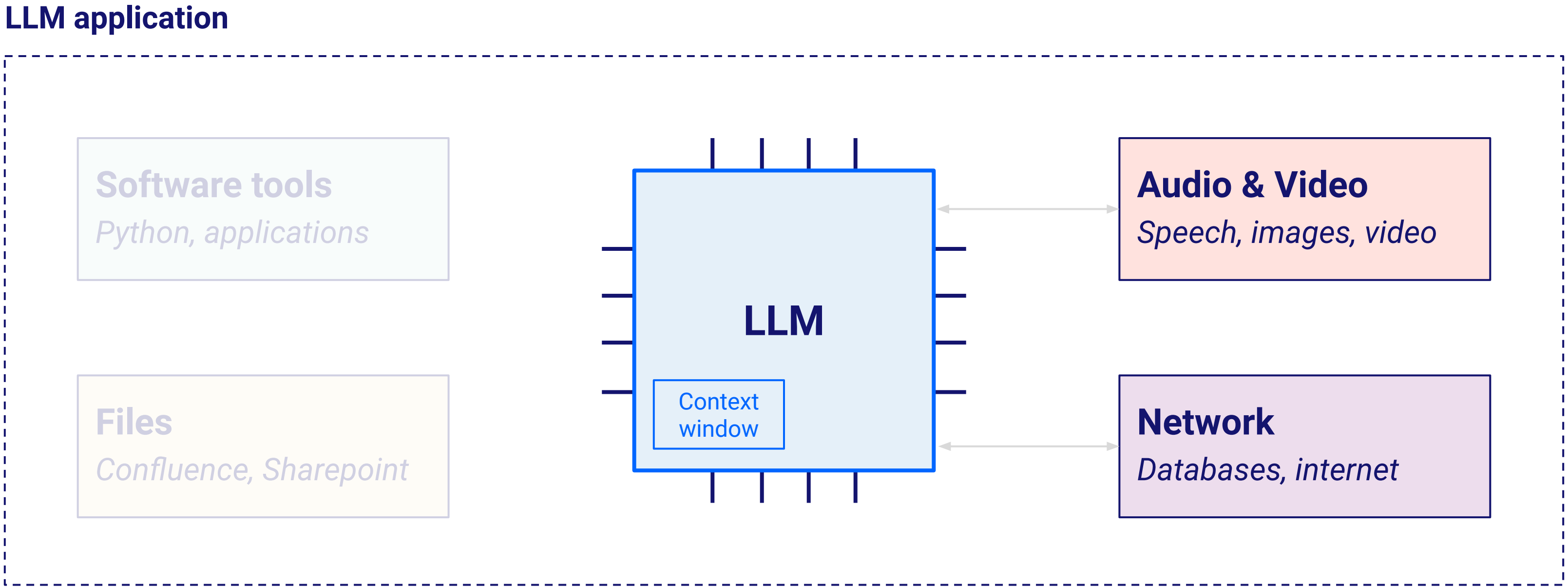
★ GPT-3 achieves a few-shot performance on MMLU of 44%: https://arxiv.org/pdf/2009.03300.pdf

[1] An automatically discovered chain-of-thought prompt generalizes to novel models and datasets: https://arxiv.org/pdf/2305.02897.pdf

[2] Language Models are Few-Shot Learners: https://arxiv.org/pdf/2005.14165.pdf

[3,4,5] Self-consistency: https://arxiv.org/pdf/2203.11171.pdf, Self-reflection: https://evjang.com/2023/03/26/self-reflection.html, Tree of Thoughts: https://arxiv.org/pdf/2305.10601.pdf

LLM app: StoryTeller podcast



LLM application

Software tools
Python, applications

Files
Confluence, Sharepoint

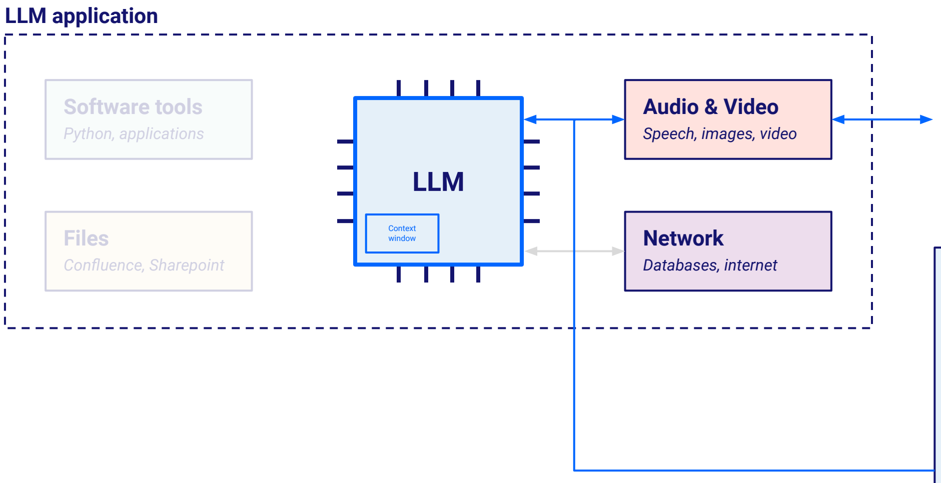
LLM

Context window

Audio & Video
Speech, images, video

Network
Databases, internet

LLM app: StoryTeller podcast



Spotify

Apple Podcasts

---Start Podcast Script---

Adam: Welcome back to Planet Money. I'm Adam Davidson, and with me as always, is Alex Blumberg. How are you doing today, Alex?

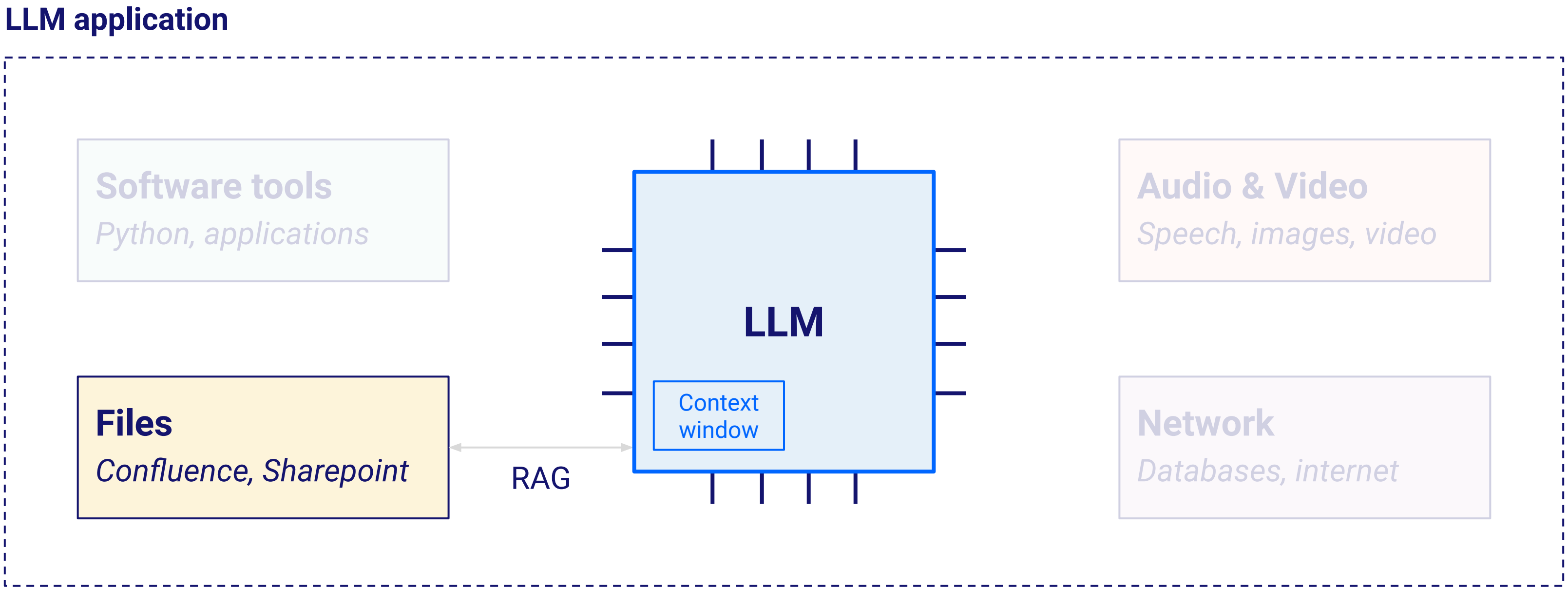
Alex: Doing great, Adam. Ready to unpack some finance stories for our listeners.

Adam: Excellent. So the mystery box we're opening today is the ING Sustainable Balanced EUR R Cap fund. Ready to dive in?

(Ambient instrumental- "intriguing" music playing)

TUE
12
14
Calendar
12.198
Mail
Photos
Camera
Settings
Finance
Weather
Clock
Navigation
Entertainment
Photos
1
X
Music
ChatGPT
9
Slack
Home
L
Lifesum
Search

LLM app: Lexicon Assistant



LLM application

Software tools
Python, applications

Audio & Video
Speech, images, video

LLM

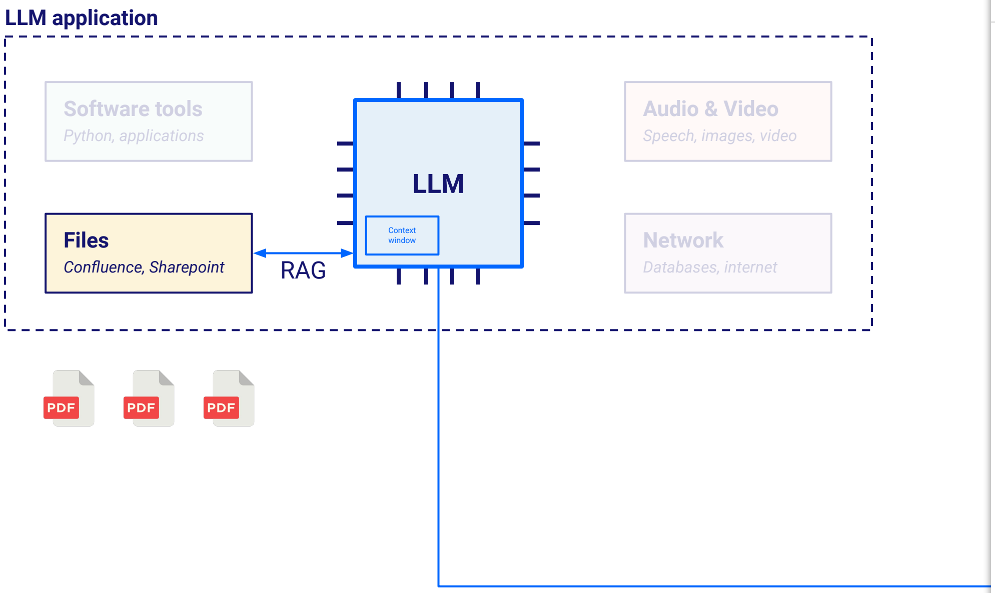
Context window

Files
Confluence, Sharepoint

RAG

Network
Databases, internet

LLM app: Lexicon Assistant



LLM application

Software tools
Python, applications

Audio & Video
Speech, images, video

LLM

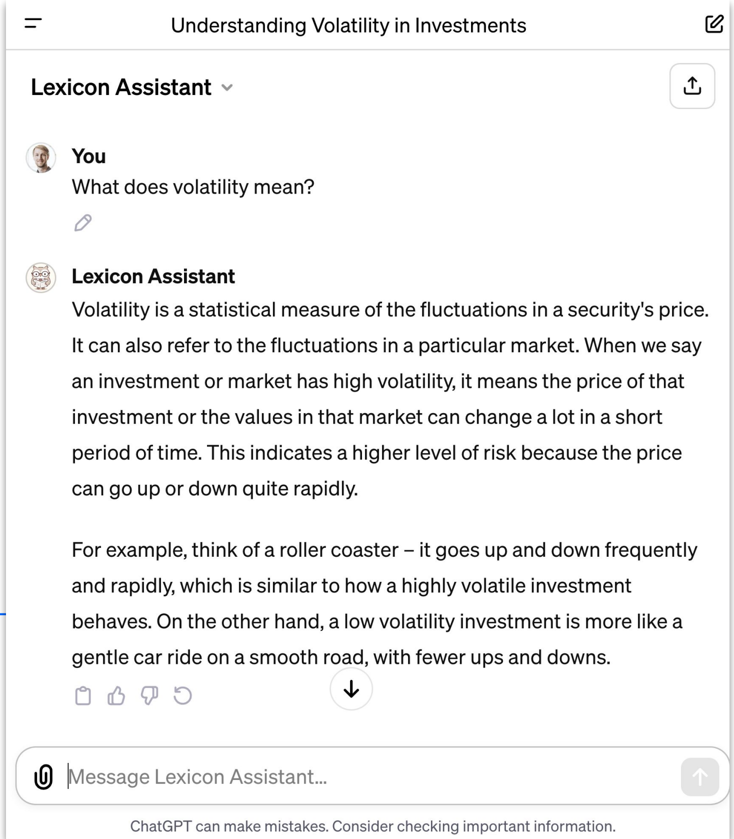
Context window

Files
Confluence, Sharepoint

RAG

Network
Databases, internet

PDF
PDF
PDF


Understanding Volatility in Investments
Lexicon Assistant
Message Lexicon Assistant...
ChatGPT can make mistakes. Consider checking important information.

You

What does volatility mean?

Lexicon Assistant

Volatility is a statistical measure of the fluctuations in a security's price. It can also refer to the fluctuations in a particular market. When we say an investment or market has high volatility, it means the price of that investment or the values in that market can change a lot in a short period of time. This indicates a higher level of risk because the price can go up or down quite rapidly.

For example, think of a roller coaster – it goes up and down frequently and rapidly, which is similar to how a highly volatile investment behaves. On the other hand, a low volatility investment is more like a gentle car ride on a smooth road, with fewer ups and downs.

Outline

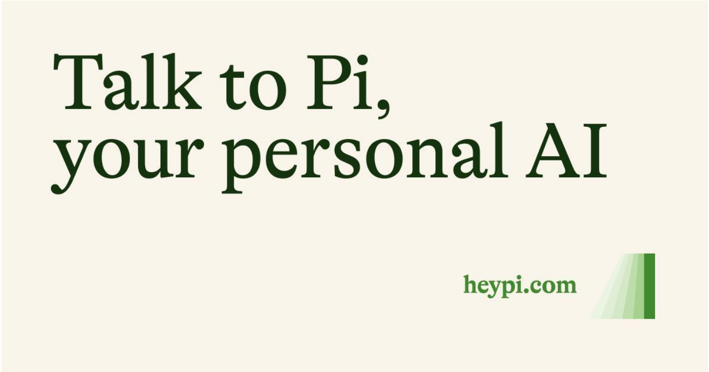
- Who are InvestSuite & Radix
- What are Large Language Models
- The magic of LLMs
- Dispelling the magic of LLMs
- Applications of LLMs
- **What does the future hold**

Radix — Copyright © RADIX.AI BVBA. All rights reserved.

fiverr.
Gabby @ggerbus
Freelance AI Copywriter
AI TOOK MY JOB
TO THE NEXT LEVEL

19:34:24:17

Personal Computer


Talk to Pi,
your personal AI

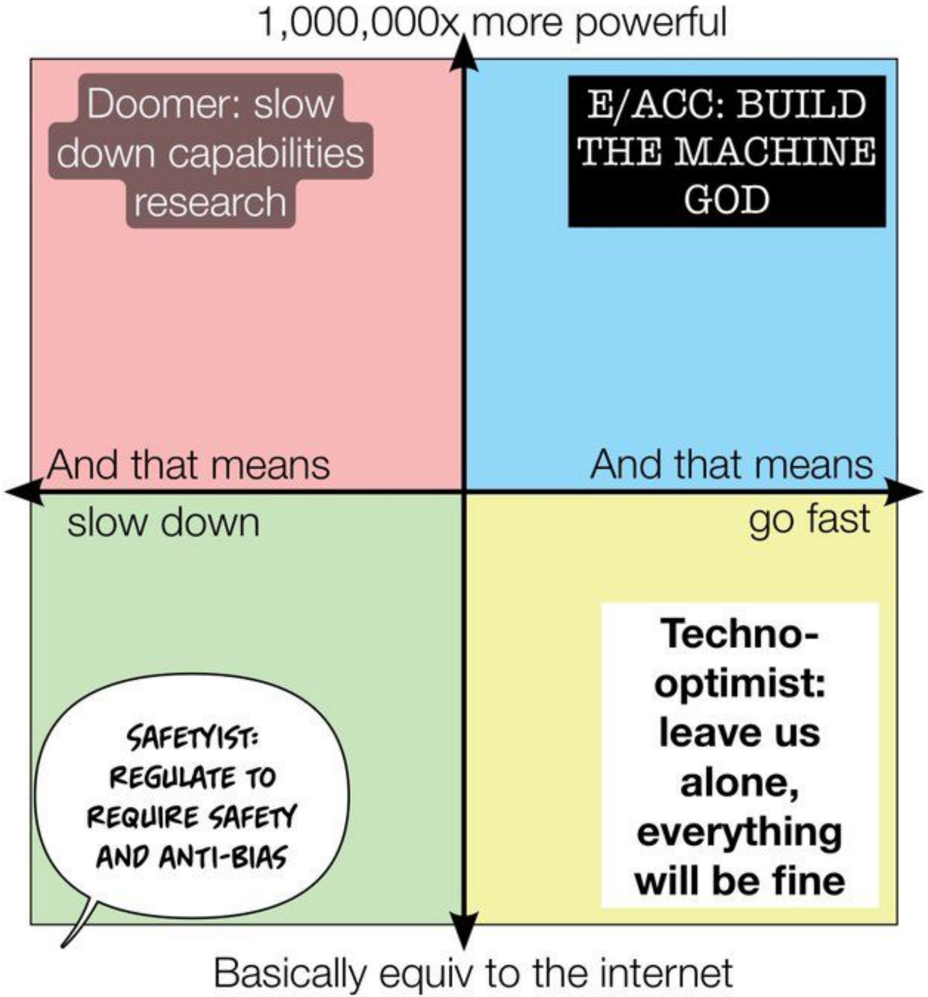
heypi.com

Personal Intelligence

Existential risk?

Doomers vs. Humanists

AI political compass


1,000,000x more powerful
Doomer: slow down capabilities research
E/ACC: BUILD THE MACHINE GOD
And that means
And that means
slow down
go fast
SAFETYIST: REGULATE TO REQUIRE SAFETY AND ANTI-BIAS
Techno-optimist: leave us alone, everything will be fine
Basically equiv to the internet

What are the H100s working on?

Inflection

Hallucination

ANTHROP\C

Alignment, Interpretability

Google DeepMind

Reasoning & planning

imbue

Reasoning & planning

OpenAI

Multi-modality

TESLA

Robotics

cohere

Retrieval

Meta

Reasoning & planning

X

‘Truth’

Radix — Copyright © RADIX.AI BVBA. All rights reserved.

AGI, when?

“Less than 10 years”[1]

Sam Altman (OpenAI)

“Within 18 months”[2]

Mustafa Suleyman (Inflection AI)

“As early as 2025, expected 2028”[3]

Shane Legg (Google DeepMind)

---

[1] Sam Altman (OpenAI) interview, 21 October 2023: https://www.youtube.com/watch?v=byYlC2cagLw
[2] Mustafa Suleyman (Inflection AI) interview, 14 November 2023: https://www.youtube.com/watch?v=VUOPCIgj-7U
[3] Shane Legg (Google DeepMind) interview, 26 October 2023: https://www.youtube.com/watch?v=Kc1atfJkiJU

AGI, when?

~~“Less than 10 years”~~[1]

Sam Altman (OpenAI)

“Within 18 months”[2]

Mustafa Suleyman (Inflection AI)

“As early as 2025, expected 2028”[3]

Shane Legg (Google DeepMind)

---

[1] Sam Altman (OpenAI) interview, 21 October 2023: https://www.youtube.com/watch?v=byYlC2cagLw

[2] Mustafa Suleyman (Inflection AI) interview, 14 November 2023: https://www.youtube.com/watch?v=VUOPCIgj-7U

[3] Shane Legg (Google DeepMind) interview, 26 October 2023: https://www.youtube.com/watch?v=Kc1atfJkiJU

Thank you!

InvestSuite

Radix