



# NUMERAI MEETUP JAPAN 2023

Stock Prediction using ChatGPT



April. 8, 2023  
UKI@blog\_uki

# 本日のプレゼン内容

## 本プレゼンの目的

- ・試してみる
- ・手順を共有する
- ・課題を抽出する
- ・儲かるよ、という報告ではない。

2部構成としました。

「ChatGPTによる株価予測」

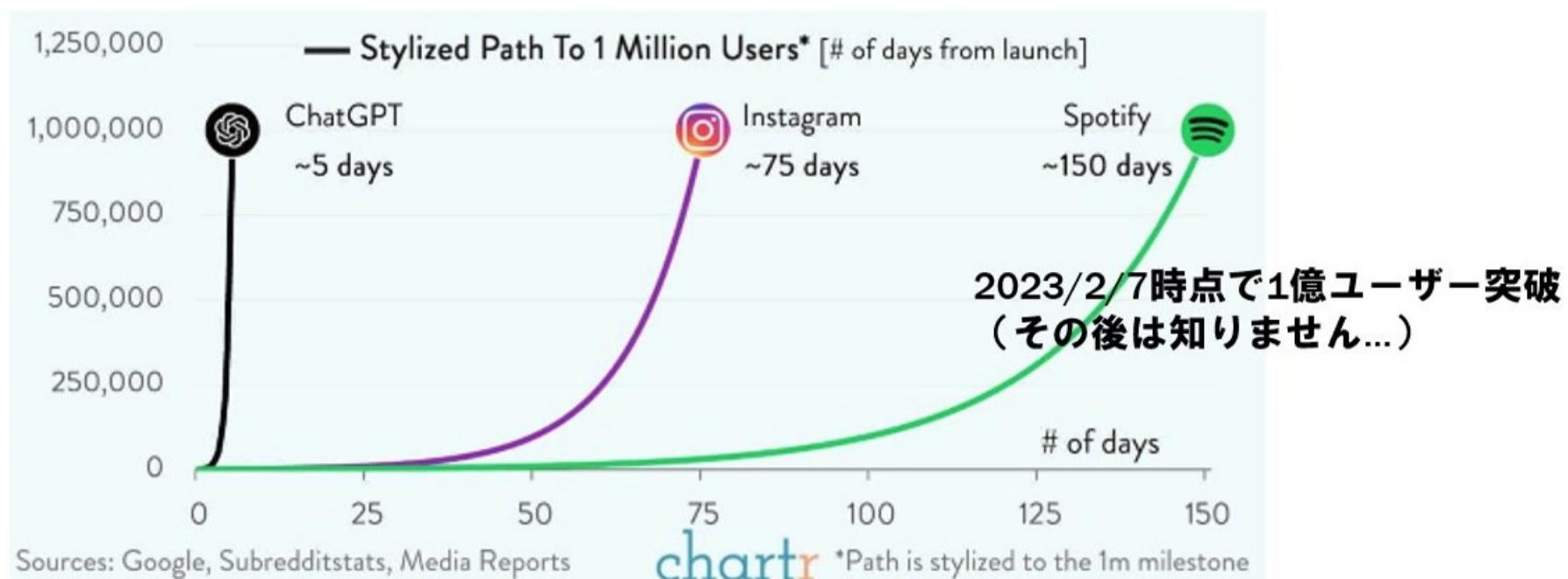
「歴史に学ぶChatGPTをはじめとするAI投資の戦略策定」

# ChatGPTによる株価予測

## ChatGPTについて

生成AI=入力に対して何かしらの出力を生成する。

- GPTとは、Generative Pre-trained Transformer
- ChatGPTはGPT3.5を搭載した大規模言語モデル（LLM）



これまでBERT等を使って四苦八苦していた自然言語処理がAPIで簡単に行える

# ChatGPTを使った株価予測フレームワーク（よくない事例）

直接聞くのはよくない。



以下の決算報告書を読んで、この企業の株価が上昇するか下落するか推測しなさい。

大和ハウス工業株式会社の四半期報告書によると、当第四半期連結累計期間において、売上高は前年同期比で増加し、経常利益や親会社による分配益も増加した。戸建住宅事業では、耐震性能や天井高仕様などの重要な変更ではなく、経営上は、マンション事業、賃貸住宅連事業を展開しており、それは新規分譲マンションの好調により、売上高と営業利益が増加した。一方、事業施設事業では物流施設関連の新規着工や安全認証に適応した施設建設の提案を強化し、建築・土木事業ともに堅調に推移している。大和ハウス工業株式会社の四半期報告書によると、物流事業では新しい物流センターがオープンし、フィットネスクラブ事業では新しい店舗がオープンしたことにより、売上高と営業利益が前年同四半期比で増加した。しかし、住宅市場では人口減少による新設住宅着工戸数の減少が見込まれ、消費増税後に市場全体が冷え込むことが懸念されている。大和ハウスグループは、将来の環境変化に備えつつ、幅広い領域の多様な収益源を活かし、計画達成を目指している。研究開発活動の金額は前年同四半期比で増加した。議決権の状況については、直前の基準日にに基づく株主名簿による記載をしている。大和ハウス工業株式会社の四半期報告書によると、同社は大量の株を保有しており、従業員持株インセンティブプランの株式や役員向け株式交付信託の株式も保有している。役員の異動ではなく、四半期連結財務諸表は規則に基づいて作成され、有限責任監査法人トーマツによる四半期レビューを受けている。四半期連結損益計算書によると、同社は営業外収益を得ており、経常利益もある。しかし、特別損失もあり、法人税等調整後の四半期純利益は非常に低い。大和ハウス工業株式会社の四半期報告書によると、四半期包括利益や保証債務、セグメント情報などが記載されている。また、連結の範囲や持分法適用の範囲に変更があったことが注目される。セグメント情報には、戸建住宅、賃貸住宅、マンション、住宅ストック、商業施設、事業施設などが含まれている。全社費用は報告セグメントに帰属しない一般管理費及び試験研究費であり、セグメント利益又はセグメント損失は四半期連結損益計算書の営業利益に調整を行っている。大和ハウス工業は、米国東部における戸建住宅開発・販売事業を展開する企業を子会社として取得し、のれんの金額に重要な変動が生じたことを報告している。のれんの増加額は、その他事業セグメントにおいて百万円である。また、四半期純利益金額は円であり、普通株式の期中平均株式数は千株である。その他、報告書には該当事項はないとされている。

**決算書を読んで  
株価が上昇するかどうか教えて**



この企業の株価が上昇するか下落するかを判断するためには、この報告書の情報だけでは不十分であり、市場の動向や他の情報も考慮する必要があります。ただし、この報告書には、大和ハウス工業株式会社が複数の事業分野で受注拡大やサービス向上に努めており、売上高と営業利益が増加していることが記載されています。また、研究開発活動の金額が増加していることや、従業員持株インセンティブプランの株式や役員向け株式交付信託の株式を保有していることが示唆されています。一方、住宅市場においては新設住宅着工戸数の減少が見込まれていることが懸念されています。以上の情報を総合すると、この企業の株価が上昇するか下落するかは不明ですが、業績の一定の好調さが示唆されていると言えます。

**テンプレ回答しか返ってこない  
・情報が不十分である  
・不明、分からない**

<使った感触>

- ・株価は予測できないと釘を刺される
- ・STEP by STEPで聞いても難しい  
(逆に誘導してしまう可能性)

そもそも株価の予測は非常に難しい。文章だけで判断できるはずがない。

# ChatGPTを使った株価予測フレームワーク（オルタナティブ）

株価を聞くのではなく、様々な情報を抽出してもらう。



以下の文章は、ある企業のトップメッセージです。この文章を読んで各年代の人人がどの程度興味を持つか、1~5で数値化しなさい。

## # 回答フォーマット

10代: 1~5

20代: 1~5

30代: 1~5

40代: 1~5

50代: 1~5

- トップメッセージを入力する**
- ・ 実際にはスクレイピングで抽出する
  - ・ GPT4以降では画像入力もできる(予定)

## # 文章

日頃より格別のご高配を賜り厚く御礼申し上げます。

4月より頭取に就任した半沢でございます。

新型コロナウイルス感染拡大の影響を受けられた皆さまに心よりお見舞い申し上げます。当行を含むMUFGグループは、経済・社会が一日も早く活力と安定を取り戻せるように、事業継続・生活確保を支える社会の礎としての責任を果たして参る所存です。

コロナ禍における健康への脅威や少子・高齢化への対応、そして、急速に進展する社会のデジタル化や脱炭素社会の実現に向けた取り組みは、私たち金融業界も含めた産業構造や社会経済に変革をもたらし、次なる大きな成長へと繋がっていきます。

全てのステークホルダーの皆様が課題を乗り越え、持続可能な未来、新たなステージへと進む時、その思いを叶えるチカラになること、そのために全力を尽くすこと、それは今後も変わらない私たちのバーバス（存在意義）です。MUFGグループでは、「世界が進むチカラになる。」というバーバスを掲げ、これを実現する新たな中期経営計画を4月よりスタートさせました。スピード重視で事業モデルの変革を進めます。そして、皆様の課題解決に貢献し、未来を切り拓く強い銀行を目指して進んで参ります。

これからも一層のご支援を賜りますよう、よろしくお願い申し上げます。

## | ごあいさつ

日頃より格別のご高配を賜り厚く御礼申し上げます。  
4月より頭取に就任した半沢でございます。

新型コロナウイルス感染拡大の影響を受けられた皆さまに心よりお見舞い申し上げます。当行を含むMUFGグループは、経済・社会が一日も早く活力と安定を取り戻せるように、事業継続・生活確保を支える社会の礎としての責任を果たして参る所存です。

コロナ禍における健康への脅威や少子・高齢化への対応、そして、急速に進展する社会のデジタル化や脱炭素社会の実現に向けた取り組みは、私たち金融業界も含めた産業構造や社会経済に変革をもたらし、次なる大きな成長へと繋がっていきます。

全てのステークホルダーの皆様が課題を乗り越え、持続可能な未来、新たなステージへと進む時、その思いを叶えるチカラになること、そのために全力を尽くすこと、それは今後も変わらない私たちのバーバス（存在意義）です。MUFGグループでは、「世界が進むチカラになる。」というバーバスを掲げ、これを実現する新たな中期経営計画を4月よりスタートさせました。スピード重視で事業モデルの変革を進めます。そして、皆様の課題解決に貢献し、未来を切り拓く強い銀行を目指して進んで参ります。

これからも一層のご支援を賜りますよう、よろしくお願い申し上げます。



株式会社三井UFJ銀行  
取締役頭取執行役員  
**半沢 淳一**



10代: 2

20代: 3

30代: 4

40代: 4

50代: 3

←JSON形式での回答も可能  
=構造化データとして直接扱える

理由：

この文章は、ある企業のトップメッセージであり、その企業の方針やビジョンについて述べられています。一般的には、若年層よりも中高年層の方が、企業の方針やビジョンに興味を持ちやすい傾向があります。しかし、この文章は新型コロナウイルス感染拡大の影響や社会の変化にも触れており、若年層にも一定の興味を持たせる要素があると考えられます。そのため、10代ではやや低めの評価となりましたが、20代以上では比較的高い評価となっています。

様々な発想を元にオルタナティブなデータを抽出し、株価との関連性を調べていく

# ChatGPTの登場が個人投資家へどのような影響を与えるか

## 従来の手法＝要素技術の塊であった

- ・テキストスクレイピング&前処理
- ・特徴量抽出（形態素解析、構文解析）
- ・数値化（極性辞書など）
- ・トピック抽出、因果チェーン検出など

## サービス1.Quick決算短信抽出



### Data Set

#### 決算短信抽出データ

提供頻度	決算短信発表毎
提供時間	決算短信発表後1時間以内
提供方法	API
情報元	QUICK

TDnetでの適時開示情報のうち「決算短信」と「業績予想の修正」から、以下の情報を抽出し提供します。

- ・セクションごとのポジネガ数
- ・セクションごとのキーワード
- ・円とドル・ユーロの想定為替レート
- ・原油、燃料油、ナフサ、金、銅、亜鉛・鉛、ニッケルの価格
- ・「継続企業の前提に関する重要事項等」または「継続企業の前提に関する注記」の記載有無

## サービス2.FINATEXT XENO FLASH（決算要約）



## サービス3.時事ドットコム 適時開示・重要事実スコア



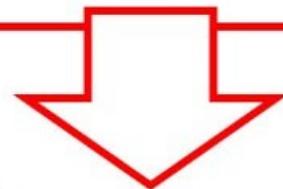
これらのサービスはChatGPTで代替可能。さらにデータを効率的に扱える時代へ

# ChatGPTへの入力データの候補

## <従来のオルタナティブデータ>

- ・衛星画像、携帯GPS、電力需要
- ・POS、クレジットデータ
- ・WEBトラフィック

- ・ニュース、適時開示
- ・WEBスクレイピング
- (経営理念、社是、表情、スピーチ)



## <今回の対象>

**適時開示系は時系列での情報取得が容易。**

- ・決算短信
- ・有価証券報告書
- ・四半期報告書、etc

参考書籍：入門オルタナティブデータ

データ種類	概要	普及度合い	主要なプレーヤー
ウェブスクレイピング	・ウェブ上の公開情報を自動収集し、投資観点で有用な情報・データベースに集約。価格.comやスカイスキャナー等の価格比較サイト等が対象にされている。	◎	YipitData、Thinknumなど
SNS	・TwitterやSansan等のSNSサイト上のコメントやソーシャルグラフ情報を活用し、消費者の「関心」をトラッキングする。	△	MarketPsych、ホットリンクなど
携帯位置情報	・携帯電話の位置情報を使って人の移動情報、滞在情報をトラッキングする。 ・感染状況との相関も高く、コロナ禍で最も活躍するデータの1つ。	○	Thasos、KDDI、クロスロケーションズなど
ウェブトラフィック	・ウェブサービスやモバイルアプリ等を主力事業としているインターネット系企業を対象に、同事業の好不調をGoogle Analyticsのような解析ツールで把握した情報を活用して分析サービスとして展開。	◎	Similar Web、App Annie、マイクロアドなど
衛星画像	・衛星画像により、穀物の生産動向や人の移動具合を把握し、先物市場の予測や小売り・テーマパーク等の好不調のトラッキング等を行う。	△	Orbital Insight、Space knowなど
新聞記事データ	・新聞記事情報やメディアニュース情報を活用しM&Aのイベント把握や、マーケットテーマの抽出、企業の取引情報抽出等を行う。	△	RavenPack、日本経済新聞、QUICKなど
消費者購買データ	・POSデータやクレジットカードデータを活用し、消費財の商品別の売上・価格の短期動向を把握、企業決算や業界動向の先読みを行う。	◎	Yodlee、Second Measure、JCB、ナウキャストなど

(出所) 株式会社ナウキャスト作成。

**ChatGPTを使って決算報告書の要約＆オルタナティブデータ抽出を行う。**

# 報告書の種類と選定

※ChatGPT3.5は最大4096トークン、ChatGPT4は最大32kトークン

	決算短信	有価証券報告書	四半期報告書
文書			
文字数	~10ページ(5000字)程度	~200ページ(~15万字)程度	~40ページ(~3万字)程度
概算トークン*	~7000トークン	~20万トークン	~4万トークン
発行頻度	4回/年	1回/年	3回/年
記載事項	主に決算内容(PL、BS) 特徴抽出する旨味が少ない	沿革、事業内容、関連会社、 従業員状況等多岐にわたる	主に事業の状況

決算短信は除外。残り2つの文書のどちらを選定するか？

# 文書の要約やデータ抽出に掛かる実行時間

※APIレスポンスはざっくり10秒/回で試算  
回答のトークン数が多いとレスポンスが長い

ユニバース	年数	文書/更新頻度		API回数	賞味実行時間
TOPIX100 100銘柄	5	有価証券報告書	1	平均15万トークン/3000=50回	2.8日
		四半期報告書	3	平均3万トークン/3000=10回	1.7日
TOPIX500 500銘柄	5	有価証券報告書	1	50	14日
		四半期報告書	3	10	8.6日
TOPIX Small1 約1000銘柄	5	有価証券報告書	1	50	28日
		四半期報告書	3	10	17日

今回は諸所の都合上、TOPIX100銘柄の四半期報告書を対象とした。

# 検証の手順

STEP1.データ収集

STEP2.文章抽出

STEP3.要約

STEP4.オルタナティブデータ抽出

STEP5.累積リターンの確認

# STEP1. データ収集

有価証券報告書、四半期報告書はEDINETから過去分も全てダウンロード可能。

<https://disclosure2.edinet-fsa.go.jp/WEEK0010.aspx>

EDINET Electronic Disclosure for Investors' Network

金融庁 閲覧

トップページ 書類詳細検索 書類全文検索 公告一覧 EDINETタクソノミ及びコードリスト ダウンロード

文字の大きさ 標準 大 English

トップメニュー

- お知らせ一覧
- よくある質問 /便利な使い方  
(2023年3月18日更新)
- 利用環境
- 操作ガイド等

システムメンテナンス情報

現在、サービス一時停止の予定はありません。

金融商品取引法等に基づき有価証券報告書等の提出期限の延長が承認されている会社一覧

開示義務違反等に係る情報をお持ちの方はこちらへ  
(ディスクロジヤー・ホットライン)

書類種別

有価証券報告書／半期報告書／四半期報告書  
 大量保有報告書  臨時報告書  その他の書類種別  
(各訂正報告書を含みます。)

提出期間

過去 1 年

重要なお知らせ

「有価証券報告書の作成・提出に際しての留意すべき事項及び有価証券報告書レビューの実施について (2023年度)」

2023年3月24日に、金融庁は、2022年度の有価証券報告書レビューの結果等を踏まえて、有価証券報告書の作成・提出に際しての留意すべき事項を公表するとともに、2023年度の有価証券報告書レビューの実施に関する公表を行いました。詳細は、以下の金融庁のページをご参照ください。  
<https://www.fsa.go.jp/news/r4/sonota/20230324-3/20230324-3.html>

「記述情報の開示の好事例集2022」の更新

全取扱い会社 2023年1月31日に公表した「記述情報の開示の好事例集(2022)」について、以下の項目を追加し、公表します。たかでお知らせいたします。

頑張って集めましょう。

## STEP2.文章抽出 ~ STEP3.要約

とにかく入力は4096トークン以内でなければならない。  
(長い回答を出力させるならもっと短く)

### <方針①>

本文を4096トークン以内に分割して、各々の文章から直接抽出を行う。  
→設問が増えると莫大な時間が掛かる。

### <方針②>

本文を4096トークン以内に分割して、いったん要約してとりまとめる。  
その後、要約した文章に対して抽出を行う。  
→一度要約すれば後の手間を減らせる（要約過程で情報喪失の危険は伴う）

- ・結局どちらの手法でも、本文中の情報が分割される恐れは残る。
- ・GPT4以降でトークン数上限は増えるので今ここで考え込んでも仕方ない。

今回は上記の方針②を採用した

# STEP2.文章抽出 ~ STEP3.要約

## ①PDF

## ②文章抽出

## ③最終要約文

提出書類 大和ハウス工業株式会社 四半期報告書 表紙 提出書類 檢査済み 文提出先 提出日 四半期報告書 金融商品取引法第 条の 第 項 開東財務局長 平成 年 月 日 四半期会計期間 第 期 第四半期 平成 年 月 日 会社名 英証名 大和ハウス工業株式会社 代表者の役職氏名 代表取締役社長 大野 直也 本店の所在の場所 大阪市北区梅田三丁目 番号 電話番号 大和 ハウス工業株式会社 代表者の役職氏名 上席執行役員 常務 山田 拓次 最寄りの連絡場所 東京都千代田区霞が関一丁目 番号 電話番号 東京 事務連絡者氏名 東京本社経理部長 中里 誠行 送り先に供する場所 大和ハウス工業株式会社 東京本社 東京都千代田区霞が関一丁目 番号 大和ハウス工業株式会社 代表者の役職氏名 名古屋市中村区平池町四丁目 番号 大和ハウス工業株式会社 横浜支店 横浜市西区みなとみらい三丁目 番号 大和ハウス工業株式会社 神戸支店 神戸市中央区元町一丁目 番号 電話番号 東京証券取引所 東京都中央区日本橋四丁目 番号 提出書類 第四半期報告書 第一部 企業情報 第 企業の概況 主要な経営指標等の推移 四次 会計期間 第 期 第四半期累計期間 第四半期累計期間 第一期 自 平成 年 月 日 至 平成 年 月 日 自 平成 年 月 日 至 平成 年 月 日 売上高 経常利益 純会計株主に帰属する四半期 当期 純利益 百万円 百万円 四半期包括利益又は包括利益 百万円 純資産額 百万円 百万円 株当たり四半期 当期 純利益 百万円 潜在株式調整後 株当たり四半期 当期 純利益 百万円 自己資本比率 % 注 当社は四半期連結財務諸表を作成しているため 提出会社の主要な経営指標等の推移については記載していません 売上高には 消費税等は含んでいません 事業の内容 当 第四半期連結累計期間において 当社グループ						
回次	提出書類	提出日	提出日	提出日	提出日	提出日
決算年度	平成20年3月期	平成20年3月	平成20年3月	平成20年3月	平成20年3月	平成20年3月
売上高	(億円)	2,007.388	2,198.119	2,190.774	2,191.860	2,192.969
粗利潤	(億円)	145.395	159.365	152.628	155.987	158.529
粗利潤率(%)		7.2%	7.2%	7.1%	7.2%	7.2%
粗利潤に繰越する 粗利潤額	(億円)	66.274	182.099	111.131	165.577	205.169
粗利潤額	(億円)	90.529	146.657	745.935	97.936	199.257
粗利潤率(%)		7.3%	7.3%	7.3%	7.3%	7.3%
粗利潤額	(億円)	736.081	952.656	1,132.071	1,181.860	1,325.981
粗利潤額	(億円)	2,371.238	2,465.946	3,821.007	3,257.893	3,555.885
粗利潤率(%)		10.6%	11.0%	11.9%	11.9%	11.9%
粗利潤に繰越する 粗利潤額	(億円)	—	—	—	115.83	304.05
粗利潤率(%)		—	—	—	—	—
自己資本比率(%)		38.9	37.0	36.6	35.9	36.4
自己資本比率(%)		5.53	11.97	11.25	9.18	16.28
自己資本比率(%)		11.89	10.97	11.34	20.24	10.91
粗利潤による キャッシュ・フロー	(億円)	144.247	161.451	178.465	216.697	205.481
粗利潤による キャッシュ・フロー	(億円)	1140.736	1248.429	1229.927	1232.447	1303.640
粗利潤による キャッシュ・フロー	(億円)	218.433	116.111	121.392	128.180	101.066
粗利潤による キャッシュ・フロー	(億円)	245.877	196.710	234.580	186.923	213.389
粗利潤による キャッシュ・フロー	(億円)	38.261	22.458	24.993	21.991	28.779
粗利潤による キャッシュ・フロー	(億円)	116.071	118.445	119.063	119.361	120.345
注: 平成20年3月期						
注1: 売上高は、消費税等を含んでいません。						
注2: 第二回から算出した粗利潤額は、各回たゞ当期粗利潤については、粗利潤額を有している粗利潤						
注3: いたたひに記載していません。						

**プロンプト**  
**あなたはプロの企業アナリストです。これから与える文章は、企業の四半期報告書の一部です。以下の注意点に従って、文章を要約しなさい。**

### # 注意点

- 与える文章はPDFファイルから抽出したものであり、意図しない文字列やスペースを含んでいます。意図しない文字列とは、表の中にある単語を抽出したものや、グラフの中にある単語を抽出したものです。このような意図しない文字列は、適切に無視して文章を要約してください。
- 要約するとき、本文中には書かれていないことを推測したり発展させたりして要約しないようにしてください。

### # 文章

## <手順>

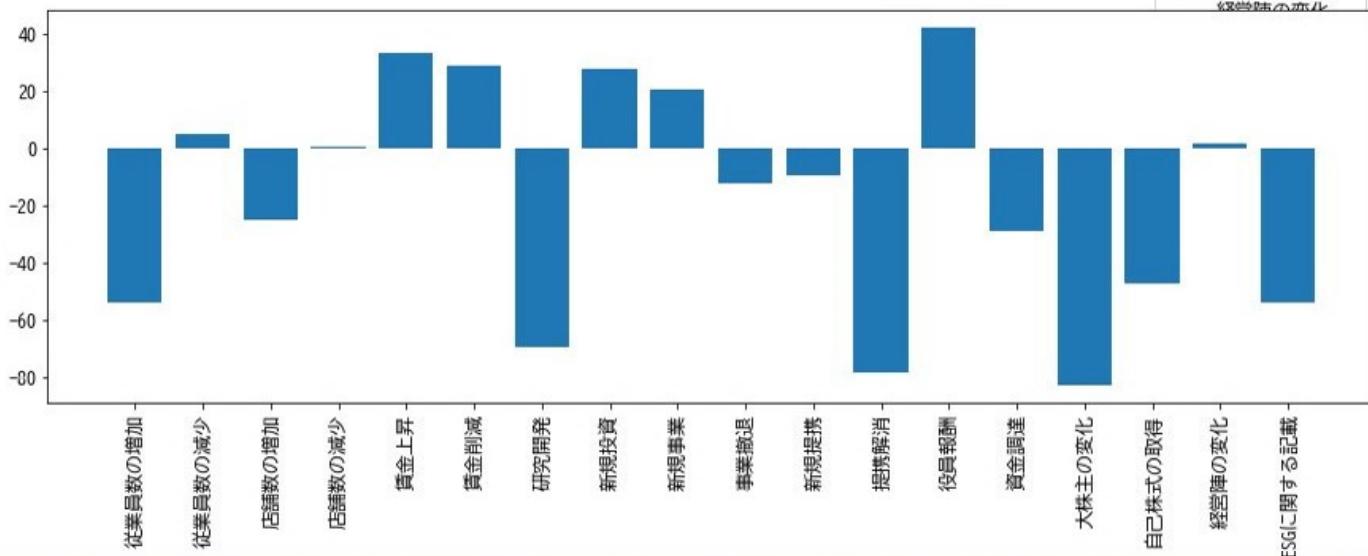
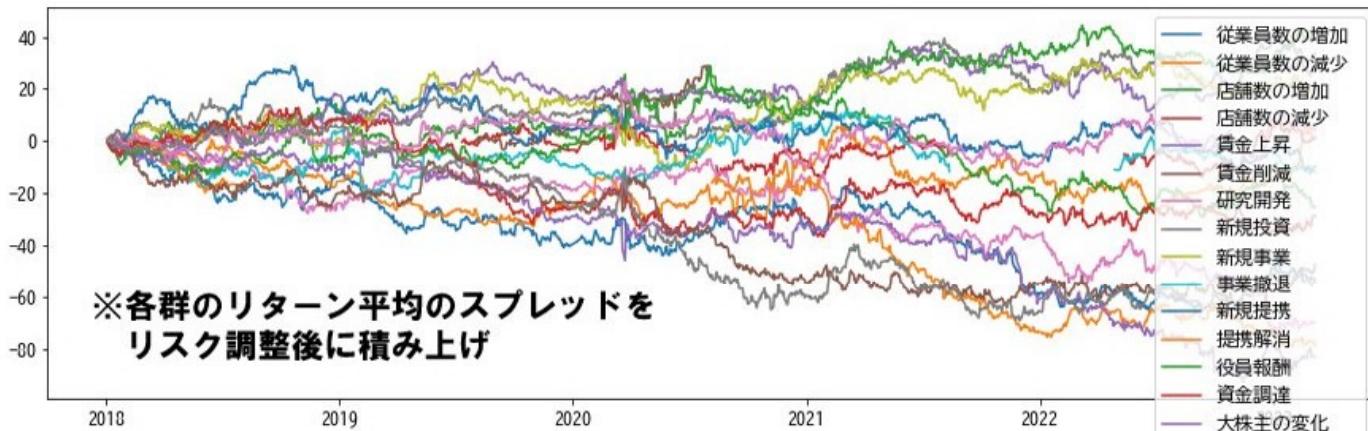
- (1) PDFの文章抽出はpdfminerなどのライブラリで簡単に抽出できる。
- (2) 文章抽出後、明らかに不要な単語が発生するのでざつと除去する。
- (3) 2000~2500文字を目途に分割する
- (4) それぞれをChatGPTで要約してもらい、最後にドッキングする。

**四半期報告書は、上記の所作で概ね扱えるようになる。**

大和ハウス工業株式会社の四半期報告書によると、当第四半期連結累計期間において、売上高は前年同期比で増加し、経常利益や親会社株主に帰属する四半期純利益も増加した。戸建住宅事業では、耐震性能や天井高仕様の改善に注力し、販売拡大を図っている。また、事業等のリスクについては重要な変更ではなく、経営上の重要な契約等の決定又は締結等もなかった。大和ハウス工業株式会社は、マンション事業、賃貸住宅事業、住宅ストック事業、商業施設事業、事業施設事業、生産施設開発事業を展開しており、それぞれの部門で受注拡大やサービス向上に努めている。特に、マンション事業では新規分譲マンションの好調な販売や子育て支援住宅の設計認定、管理事業におけるサービス充実により、売上高と営業利益が増加した。一方、事業施設事業では物流施設開発の新規着工や安全認証に適応した施設建設の提案を強化し、建築・土木事業ともに堅調に推移している。大和ハウス工業株式会社の四半期報告書によると、物流事業では新しい物流センターがオープンし、フィットネスクラブ事業では新しい店舗がオープンしたことにより、売上高と営業利益が前年同四半期比で増加した。しかし、住宅市場では人口減少による新設住宅着工戸数の減少が見込まれ、消費増税後に市場全体が冷え込むことが懸念されている。大和ハウスグループは、将来の環境変化に備えつつ、幅広い領域の多様な収益源を活かし、計画達成を目指している。研究開発活動の金額は前年同四半期比で増加した。議決権の状況については、直前の基準日に基づく株主名簿による記載をしている。大和ハウス工業株式会社の四半期報告書によると、同社は大量の株を保有しており、従業員持株インセンティブプランの株式や役員向け株式交付信託の株式も保有している。役員の異動はなく、四半期連結財務諸表は規則に基づいて作成され、有限責任監査法人トーマツによる四半期レビューを受けている。四半期連結損益計算書によると、同社は営業外収益を得ており、経常利益もある。しかし、特別損失もあり、法人税等調整後の四半期純利益は非常に低い。大和ハウス工業株式会社の四半期報告書によると、四半期包括利益や保証債務、セグメント情報などが記載されている。また、連結の範囲や持分法適用の範囲に変更があったことが注目される。セグメント情報には、戸建住宅、賃貸住宅、マンション、住宅ストック、商業施設、事業施設などが含まれている。全社費用は報告セグメントに帰属しない一般管理費及び試験研究費であり、セグメント利益又はセグメント損失は四半期連結損益計算書の営業利益に調整を行っている。大和ハウス工業は、米国東部における戸建住宅開発・販売事業を展開する企業を子会社として取得し、のれんの金額に重要な変動が生じたことを報告している。のれんの増加額は、その他事業セグメントにおいて百万円である。また、四半期純利益金額は円であり、普通株式の期中平均株式数は千株である。その他、報告書には該当事項はないとされている。

## STEP4.オルタナティブデータ抽出（客観的）～STEP5.リターン確認

- 記載があるかどうかを問う（既存の構造化データに含まれていないものを設問）
- ChatGPTの主觀は介在させない。



有意とは言えないが定性的に理解できるものもある

### プロンプト

あなたはプロの企業アナリストです。これから与える文章は、企業の四半期報告書の要約です。以下の項目について記載があれば1、記載がなければ0とし、JSON形式で出力しなさい。

### # 注意点

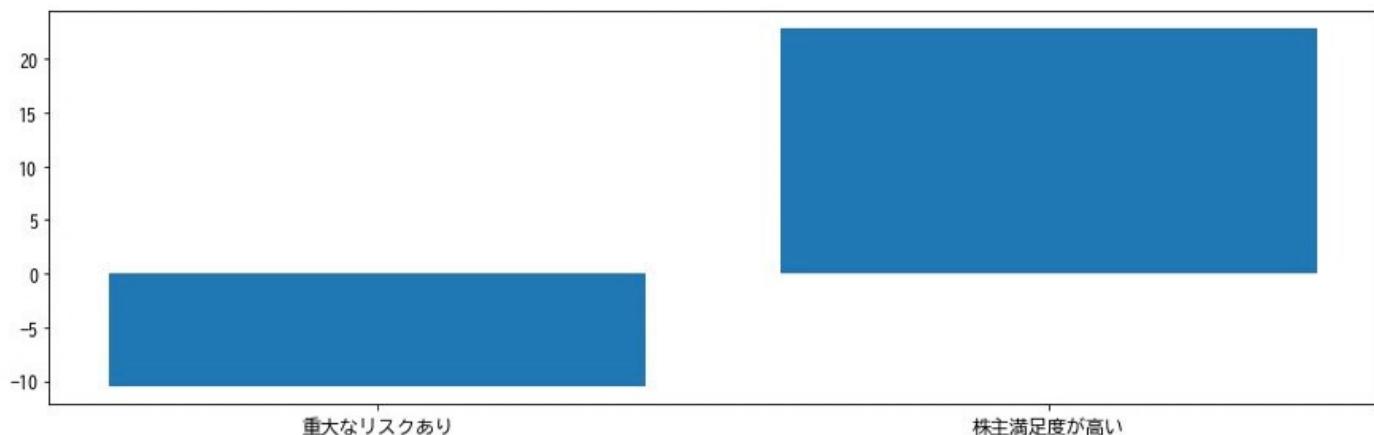
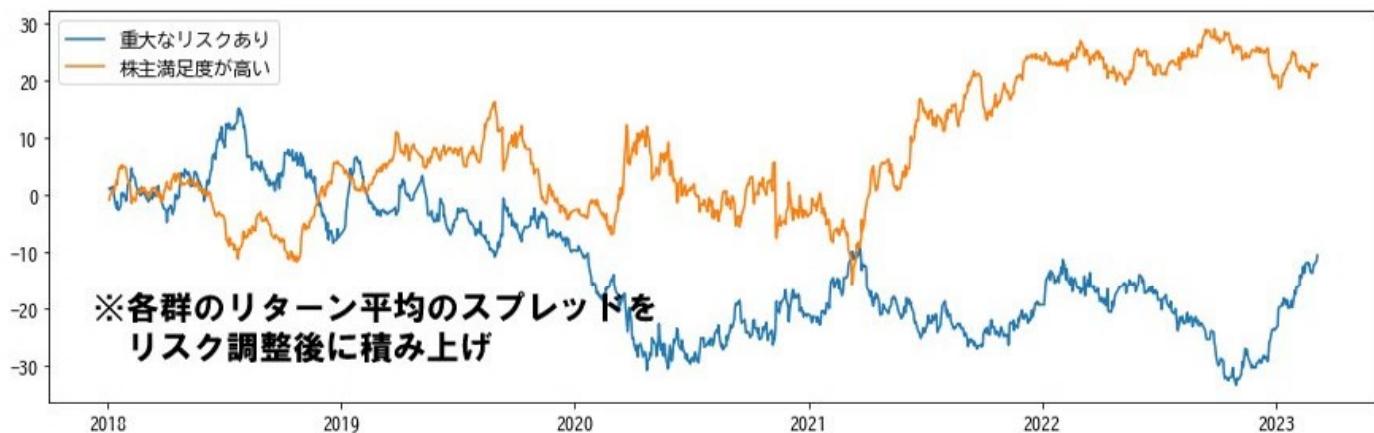
- 出力は項目名と数値のみをJSON形式で返してください。理由などの余分な記載は不要です。

### # 項目

- 働員数の増加
- 働員数の減少
- 店舗数の増加
- 店舗数の減少
- 貸金上昇
- 貸金削減
- 研究開発
- 新規投資
- 新規事業
- 事業撤退
- 新規提携
- 提携解消
- 役員報酬
- 資金調達
- 大株主の変化
- 自己株式の取得
- ESGに関する記載

## STEP4.オルタナティブデータ抽出（主観的）～STEP5.リターン確認

- ChatGPTがどう考えるか問う。
- 思考の観点は一応与えておく。



### プロンプト(1)

あなたはプロの企業アナリストです。これから与える文章は、企業の四半期報告書の要約です。以下の観点から最終的にこの企業に重大なリスクが発生しているかどうか判断しなさい。重大なリスク発生している場合は1、発生していない場合は0とし、結果の数値のみを出力しなさい。

### # 観点

- 企業を取り巻く環境に重大な変化がないか
- 不祥事などが発生していないか
- 訴訟を起こされていないか
- 関連会社の倒産はないか
- デモやストライキが発生していないか
- 海外展開が失敗していないか
- 災害によって損害が発生していないか
- 機密情報の漏洩はないか

### プロンプト(2)

あなたは株主です。これから与える文章は、企業の四半期報告書の要約です。以下の観点から最終的に株主としての満足度を5段階評価で判定しなさい。とても満足な場合は5、全く満足できない場合は1とし、結果の数値のみを出力しなさい。

### # 観点

- 業績の推移
- 新事業の展開
- 企業を取り巻く環境変化
- 内在するリスク
- 将来への展望
- 株主への誠実さ
- 株主への還元

ちゃんと考察してくるが、こちらも有意とは言えない

# ChatGPTを使った株価予測 ここまでまとめ

## 取り急ぎ触った感触

- ・とにかく楽に抽出できる。低成本。誰でもできる。
- ・客観的なデータの抽出はワークするイメージ。
- ・主観的な質問をぶつけるのはマズイ気がする（一貫性がない）
- ・プロンプトエンジニアリングで精度を上げようとする試みは失敗する可能性大。  
(定量的な正解がないため、単にオーバーフィットするだけ)
- ・総じて、既存の戦略を補う目的であれば概ね良好に使えるものと考える。

## Further Work

- ・データの拡張（有価証券報告書やそれ以外のデータソース）
- ・もっと大きなトークン数での試行（GPT4に期待）
- ・ファインチューンしたときの性能向上の余地（参考：BloombergGPT）

# 歴史に学ぶ ChatGPTをはじめとする AI投資の戦略策定

## <お断り>

- ・筆者自身は金融の業界人ではありません。
- ・筆者は個人投資家として、2016年から努めてAI投資に関する情報収集を行ってきました。
- ・本内容はそれらの情報に基づく、ごく主観的な展望となります。
- ・プロフェッショナルの方々からのご指摘・ご意見は是非とも参考にしたい所存です。  
ご意見等、賜りますようお願い申し上げます。

# AI投資・AI運用の歴史

界隈	~2012	...	2016	2017	2018	2019	2020	2021	2022	2023
ML技術	●AlexNet (2012)		●Alpha-Go	●Transformer					●Stable Diffusion	●ChatGPT
アカデミック(SIG-Fin)			●AIを用いた金融政策予測	●FXのトラッキング解析	●マルチタスク学習		●GPS情報によるナウキャスト	●MLによる経済予測モデル	●Trader-Company法	
金融コンペ			●Winton	●Two Sigma1	●Two Sigma2		●Jane Street	●Ubiquant	●Optiver	●JQuants
投信／ロボアド			<p style="text-align: center;"><b>第一次AI投資ブーム</b></p> <ul style="list-style-type: none"> <li>●ウェルスナビ ロボアド最大手</li> <li>●Yjam (PayPay投信AIプラス)</li> <li>●GSビッグデータ</li> <li>●AI日本株式オープン</li> </ul> <p style="text-align: center;">初期の投信</p>			●FOLIO ROBO	●ウェルスティング	後発のロボアド組	<p style="text-align: center;"><b>第二次AI投資ブーム?</b></p> <ul style="list-style-type: none"> <li>●DAIWA AI セレクト</li> <li>●カブコム信用ロボアド</li> <li>●Alpaca Robo</li> </ul>	
AI特化ヘッジファンド	<ul style="list-style-type: none"> <li>●Rebellion Research (2007)</li> <li>●Sentient Technologies (2007)</li> <li>●Cerebellum Capital (2009)</li> <li>●Aidyia (2011)</li> <li>●Derwent Capital (2011)</li> <li>●Quantopian (2011)</li> </ul>	<ul style="list-style-type: none"> <li>●Numerai</li> <li>●Simplex Asset Management</li> <li>●KFL Capital (2015)</li> </ul>	<p>目立ったAIファンダが観測されない</p>							

Alpha碁でAIに注目が集まった2016年の状況と酷似している印象を受ける。

# ロボアド・投信系の採用技術の変遷

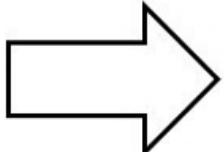
界隈	~2012	...	2016	2017	2018	2019	2020	2021	2022	2023
投信／ ロボアド			<ul style="list-style-type: none"> <li>●ウェルスナビ ロボアド最大手</li> </ul>	<ul style="list-style-type: none"> <li>●Yjam (PayPay投信AIプラス)</li> <li>●GSビッグデータ</li> <li>●AI日本株式オープン</li> </ul>	<p style="text-align: center;">初期の投信</p>		<ul style="list-style-type: none"> <li>●FOLIO ROBO</li> </ul>	<ul style="list-style-type: none"> <li>●ウェルスティング</li> </ul>		

- ・ウェルスナビ→ETFによるアセットアロケーション。預かり資産7000億。  
利回りはTOPIXと同程度。全自动を謳うが複雑なモデルを使っている印象は薄い。  
=プロモーションによる勝ち組

- ・初期の投信でTOPIXをアウトパフォームしているものは少ない。  
逆に後発組はTOPIXをアウトパフォームするものが多い。

<初期の投信が採用した技術やデータ>

- パワーワード：ビッグデータ解析
- ・決算短信、ニュース
- ・アナリスト予測
- ・特許情報
- ・Yahoo掲示板
- ・気象データ
- ・検索ワード



<後発組が採用した技術やデータ>

- ・マーケットデータ、コモディティ、マクロデータ(FOLIO)
- ・マルチファクター分析（スマートベータ）(ウェルスティング)
- ・決算発表データ、直後の値動きデータ (DAIWA AIセレクト)
- ・データ分析によるパターンマッチング (カブコム)
- ・マーケットデータによるディープラーニング (Alpaca Robo)

プロセス管理のノウハウは十分に溜まった印象

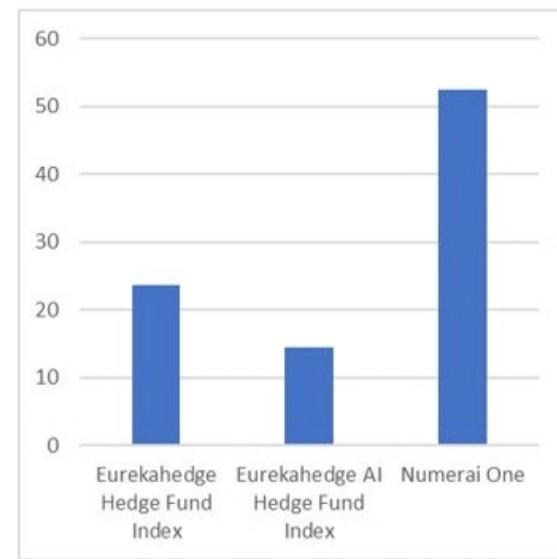
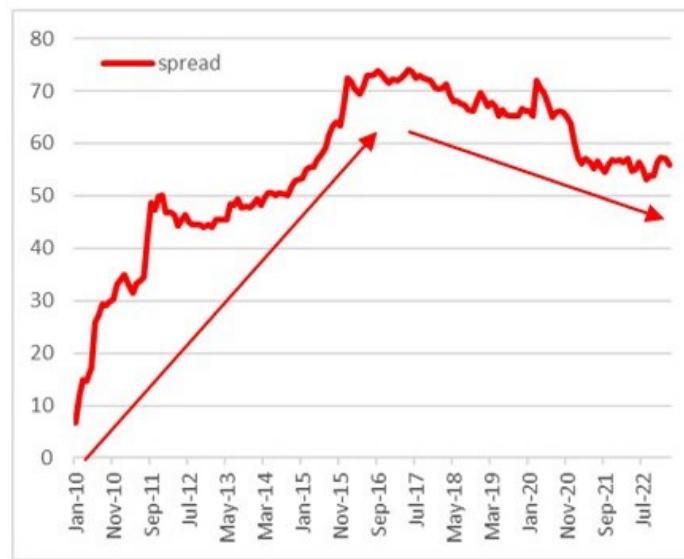
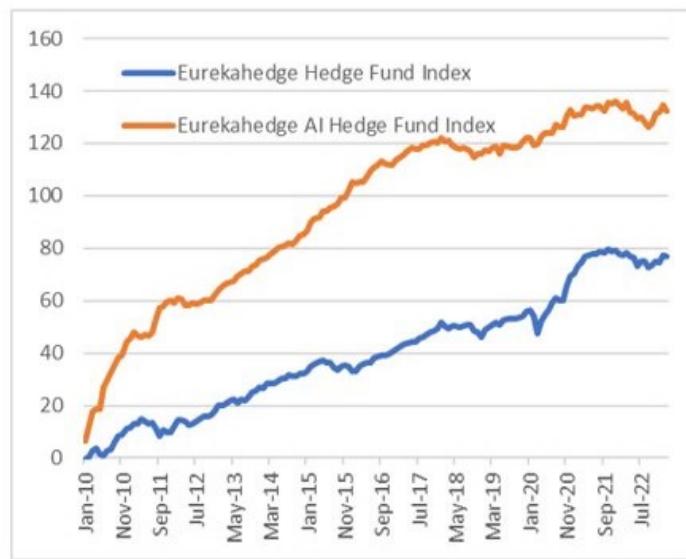
参考：「マーケットAI開発プロジェクトにおけるプロジェクトマネジメントのエッセンス」

**ビッグデータを扱うよりも丁寧なマーケットデータ分析が重要だったという印象**

# AIヘッジファンドのパフォーマンスについて

界隈	~2012	...	2016	2017	2018	2019	2020	2021	2022	2023
AI特化 ヘッジ ファンド	● Rebellion Research (2007) ● Sentient Technologies (2007) ● Cerebellum Capital (2009) ● Aidyia (2011) ● Derwent Capital (2011) ● Quantopian (2011)	● Numerai ● Simplex Asset Management ● KFL Capital (2015)								目立ったAIファンドが観測されない

- AIヘッジファンドは当初は好調であったが、2016以降は苦しんでいる。
- Numeraiは好調をキープ



Cumulative Return (Sep 19 - Feb 23)

競争が激化している可能性。AIを使うのは大前提だが戦略の選択が重要。

# AI投資に関する見立て 答え合わせ

見立て		Answer
ランダムフォレストを使って株価騰落を的中率68%で予測できる? (2016年当時のニュースヘッドライン)	No	やはり無理であった。
AIがファンドマネージャーにとって代わる? (2016年当時のニュースヘッドライン)	No	一方でAIはなくてはならない重要なツールとなつた(使っていないところはない)。
ディープラーニングが抽出する特徴は、通常のテクニカル指標と同様のものであり、有用でないのでは? (2017年に自身のブログにて言及)	No	実際には生のデータに対して絶大な効果を発揮した。リサーチ現場で使われる機械学習はほぼDLとなつた。一方でテーブルデータのコンペでは、まだまだGBMが根強い人気を誇る。
AIの本質的な活用用途は、モデリングではなくサンプリングにあるのでは。複雑なモデルはデータ収集コストの低減に利用し、予測モデルはシンプルなほうがよいのでは? (2017年に自身のブログにて言及)	No	実際にはモデリングでも大きな力を発揮した。
アルファ碁によって注目が集まつたせいで、AIを活用した投資手法の寿命は短くなるのでは? (2016年に自身のブログにて言及)	Possibly Yes	断言はできないが、2016年以降、AIヘッジファンドのパフォーマンスは通常のヘッジファンドに対して確かに悪くなつてゐる。

ChatGPTを使ったAI投資の戦略策定はどのように行うべきか?

# AI投資に関する見立て 答え合わせ

見立て		Answer
ランダムフォレストを使って株価騰落を的中率68%で予測できる? (2016年当時のニュースヘッドライン)	No	やはり無理であった。
AIがファンドマネージャーにとって代わる? (2016年当時のニュースヘッドライン)	No	一方でAIはなくてはならない重要なツールとなつた(使っていないところはない)。
ディープラーニングが抽出する特徴は、通常のテクニカル指標と同様のものであり、有用でないのでは? (2017年に自身のブログにて言及)	No	実際には生のデータに対して絶大な効果を発揮した。リサーチ現場で使われる機械学習はほぼDLとなつた。一方でテーブルデータのコンペでは、まだまだGBMが根強い人気を誇る。
AIの本質的な活用用途は、モデリングではなくサンプリングにあるのでは。複雑なモデルはデータ収集コストの低減に利用し、予測モデルはシンプルなほうがよいのでは? (2017年に自身のブログにて言及)	No	実際にはモデリングでも大きな力を発揮した。
アルファ碁によって注目が集まつたせいで、AIを活用した投資手法の寿命は短くなるのでは? (2016年に自身のブログにて言及)	Possibly Yes	断言はできないが、2016年以降、AIヘッジファンドのパフォーマンスは通常のヘッジファンドに対して確かに悪くなつてゐる。

ChatGPTを使ったAI投資の戦略策定はどのように行うべきか?

## まとめ

- ChatGPTを使った特徴抽出は、それ自体はコスト対効果が高く、既存の戦略を補う目的であれば概ね良好に使えるものと考える。
- テキスト分析を主力としたファンドのパフォーマンスは必ずこれから下がる。ChatGPTによるテキスト分析に完全に依存することは避けるべきである。
- 今回はChatGPTを使ったシンプルな戦略について検討したが、これから高いパフォーマンスを維持するためには、LLMからインスピアイアされた画期的なアイディアを探すべきである。
- 2016年にクラウドソーシングファンドという画期的な運用スタイルを編み出したNumeraiのように※。

※当時はビッグデータ解析やDLを使ったマーケットマイクロストラクチャのモデリングが主流であり、このような形態を取ったファンドはいなかった\*\*\*。

※※厳密にいうとQuantopianが存在したが、知財的な観点から有用なモデルを集めることができなかつたと思われる。

ご清聴ありがとうございました