TOWARDS PROBABILISTIC ARTIFICIAL INTELLIGENCE IN BEHAVIORIAL SCIENCES

GIACINTO PAOLO SAGGESE*, PAUL SMITH*, AND WHOEVER ELSE HELPED*

Abstract. This is draft v0.1.

Contents

1. Introduction

This paper proposes an approach to forecast variables in behavioral sciences (e.g., economics and finance, political science, business management, psychology) combining several emerging technologies in a novel way:

(1) Deep-learning based AI, in particular large language models (LLMs)
(2) Knowledge graphs (KGs)
(3) Bayesian inference for probabilistic reasoning
(4) High-performance time-series streaming computing

The proposed approach allows to deal with most of the problems plaguing forecasting in behavioral sciences such as:

- Small and noisy datasets
- Need for interpretability of forecasts

Date: April 3, 2024.
* Authors listed alphabetically.

- Need for estimation of confidence in prediction
- Complex, time-varying relationships between variables

It also allows to deal with problems of:

- LLM hallucinations
- Lack of fit (overfit or underfit) in machine learning models

2. A possible definition of AGI in behavioral sciences

We define AGI as a computer system able to perform tasks better than a human can do.

Behavioral (or social) science is devoted to the study of societies and the relationships among individuals within those societies. It encompasses a wide array of disciplines, including economics, management and business science, psychology and political sciences.

Our approach to developing AGI, with an initial focus on behavioral sciences (such as economics), reflects a strategic pathway towards achieving broader artificial general intelligence capabilities. Here's an analysis of your approach and its implications:

- **Domain-Specific Expertise as a Foundation for AGI**: Focusing on branches of behavioral sciences initially leverages the concept that substantial domain-specific knowledge is essential for creating super-human performance. E.g., finance, with its complexity and data-rich environment, provides a fertile ground for developing and refining AI algorithms. This sector demands precision, adaptability, and the ability to process vast amounts of data – attributes that are foundational for AGI.
- **Stepping Stone Approach**: By mastering one domain, you establish a methodology and technological base that can be adapted and expanded to other domains. This step-by-step approach is pragmatic, reducing the initial complexity that would come with attempting to create a multi-domain AGI from the start. Each domain mastered adds another layer of complexity and capability to the AGI, gradually moving towards true general intelligence.
- **Integration of Diverse Data Sources and Techniques**: The ability to integrate and analyze data from various sources using advanced modeling techniques is crucial. This includes real-time processing of global market data, news, economic indicators, and even sociopolitical events. Such integrative capacity is essential for AGI, as it mirrors the multifaceted nature of human intelligence.
- **Addressing the Certainty and Hallucination Problem**: In AI, hallucination refers to generating or inferring information that isn't supported by data. An essential feature of our AGI system would be its ability to estimate and report the level of certainty in its predictions and decisions. This is a significant challenge in AI development, as it requires the system to not only analyze data but also understand and communicate the limits of its knowledge and the confidence in its conclusions.
- **Ethical and Practical Implications**: As your AGI system advances, it's important to consider ethical implications, particularly in sensitive areas like economics, finance, and other behavioral sciences. Issues such as data privacy, security, and the potential impact on employment and economic systems are vital considerations. Additionally, the system's decisions and predictions should be transparent and explainable, especially when they significantly impact financial markets or individual investments.

Questions should be answered estimating a measure of certainty (this is related to the problem of hallucination)

The goal is a system that can automatically answer questions with accuracy superior to humans like:

- Predict the price of BTC in 7 minutes using fundamental info and news
- Predict the volatility of oil price in one month given the current political situation

- Predict the likelihood of an house crisis using only data available strictly before 2007
- Predict the unemployment rate this month?
- Predict how many Oscars "Dune 2" will win in 2024
- Predict the probability of a foldable IPhone by end of 2024?
- Predict the prescription increase this quarter of Ozempic
- Predict the top artist on Spotify this year
- Predict the highest temperature in Austin today
- Predict the fraud attitude of player XYZ given his betting history
- Estimate the effect of increasing 10% my marketing spending on the demand for my product XYZ
- What would be the global economic impact of a sudden 2% increase in the U.S. Federal Reserve interest rate today?
- What are the predicted movements of major currencies in the forex market over the next quarter, based on current geopolitical situations, trade relations, and economic policies?
- What will the rate of inflation in the next year in Italy? The current answer from ChatGPT is:

  "'Predicting the exact rate of inflation for the next year in any country, including Italy, is a complex task that involves analyzing a multitude of economic indicators, current monetary policies, global economic trends, and unforeseen events. As of my last update in April 2023, I don't have real-time data or the ability to predict future economic conditions."'
- Can you predict the risk of a specific stock, such as Apple, in the next 15 minutes based on real-time market data, recent company news, and advanced predictive analytics?

  The current answer from ChatGPT is: "'You are an expert economist. Can you predict the risk of a specific stock, such as Apple, in the next 15 minutes based on real-time market data, recent company news, and advanced predictive analytics? Answer in 30 words"'

  ChatGPT: "'As an AI, I don't have real-time market data or the ability to predict short-term stock movements. Accurate 15-minute forecasts require live data and are highly speculative due to market volatility."'

2.1. **The need for incorporating certainty.** Implementing a measure of certainty in AGI responses, particularly in a complex field like finance, is therefore not just a technical feature, but a fundamental aspect that enhances the utility, reliability, and ethical use of the system.

Incorporating a measure of certainty in the responses of an AGI, especially in the context of finance, is crucial for several reasons:

- **Reducing the Risk of Hallucination**: AI systems, including AGIs, can sometimes "hallucinate", meaning they generate responses or predictions that are not grounded in the data they have processed. By quantifying the certainty of its answers, an AGI can indicate the reliability of its predictions, helping users differentiate between high-confidence insights and those that are more speculative.
- **Enhancing Decision-Making**: In finance, decisions often hinge on the level of risk and uncertainty. An AGI that can estimate and communicate the degree of certainty in its analysis provides invaluable information for risk assessment. This allows users to make more informed decisions, weighing the potential risks and rewards more accurately.
- **Building Trust**: A system that acknowledges the limits of its knowledge and provides certainty estimates can build greater trust with its users. In fields like finance, where decisions can have significant consequences, trust in the system's output is essential.
- **Dynamic Learning and Improvement**: By quantifying certainty, the AGI can also identify areas where its models may need improvement. Lower certainty in certain types of

predictions can signal the need for additional data, refinement of models, or reevaluation of the algorithms used.

- **Managing Complex Systems**: Finance is a complex, dynamic system influenced by a multitude of factors. Providing a measure of certainty helps in understanding the impact of various elements and their interplay, acknowledging that some aspects of the financial world are inherently unpredictable.
- **Ethical and Responsible AI**: This approach aligns with ethical AI practices, where transparency and accountability are key. Users should be aware of both the capabilities and limitations of the AGI, and a measure of certainty facilitates this understanding.

2.2. **Why focusing on behavioral sciences?** The application of AGI in behavioral sciences is not only a test of its predictive capabilities but also an exploration into understanding complex, human-centric systems. The success in these fields can have far-reaching implications, both in terms of technological advancement and societal benefits.

The application

Building Artificial General Intelligence (AGI) in the field of behavioral sciences economics is meaningful for several reasons:

- **Limited Success of Scientific Method**: The scientific method faces challenges in behavioral sciences compared to natural sciences because human behavior and social phenomena are complex, variable, and influenced by cultural, historical, and subjective factors that are difficult to control and predict. While natural sciences can often rely on controlled, repeatable experiments to test hypotheses, the behavioral sciences deal with phenomena that cannot always be replicated in controlled environments, leading to a reliance on observational, qualitative, and interpretive research methods that may not fit the strict empirical mold of the scientific method.
- **Nature of Findings**: The findings in behavioral sciences are often context-dependent and may not be universally applicable due to the diversity of human experience and the influence of culture, history, and environment. These disciplines typically deal with phenomena that are more subjective and interpretive.
- **Understanding Complex Systems**: Finance and economics are intricate fields characterized by non-linear, interdependent variables. AGI systems could offer advanced understanding of these complexities, allowing for better predictions and insights. This is particularly valuable as financial systems are influenced by a myriad of factors including human behavior, market trends, political climates, and global events.
- **Dynamic Prediction Capabilities**: AGI in finance isn't just about static analysis but involves dynamic prediction in real-time. Financial markets are constantly evolving, and an AGI system can adapt to these changes, offering predictions and analyses that reflect current realities. This dynamic environment is a robust testing ground for AGI capabilities.
- **Managing Uncertainty and Noise**: Financial data is often noisy and uncertain. An AGI system's ability to sift through this noise and make accurate predictions despite uncertainty is a testament to its understanding of complex, real-world environments. This capability could significantly reduce risks and improve decision-making processes.
- **Immediate Practical Benefits**: Enhancements in economic and financial decision-making directly translate to benefits for the human race. Better financial predictions and economic models can lead to more stable economies, improved allocation of resources, reduced risks of financial crises, and overall economic growth. This can improve living standards and contribute to societal well-being.
- **Understanding Human Nature**: Since finance and economics are deeply intertwined with human behavior, an AGI capable of navigating these fields would necessarily gain

insights into human psychology and behavior. This is crucial for building AGI that truly understands and interacts effectively with human-centric environments.
- **Ethical and Societal Implications**: Working in these domains also necessitates dealing with ethical considerations, such as privacy, security, and fairness. Successfully navigating these issues in the realm of finance and economics can set precedents for AGI applications in other areas, leading to more responsible and ethical AI development.

2.3. **The problem with applying machine learning in economics and finance.** Applying machine learning in economics and finance presents several challenges due to the unique nature of these fields, besides the traditional problems of overfitting (i.e., fitting the noise in the training data instead of capturing the underlying true dynamics, leading to poor generalization) and underfitting (i.e., oversimplified models may fail to capture the complexities of financial data, resulting in inaccurate predictions)

Some specific reasons that make machine learning difficult to apply in economics and finance are:

(1) **Data Quality and Availability**:
- **Noise in Data**: Economic and financial data are often noisy and non-stationary. Market sentiments, geopolitical events, and economic policies can rapidly change, affecting the quality and relevance of data.
- **Limited Access**: High-quality, granular data can be expensive or restricted, limiting the scope of analysis. Proprietary data sources and privacy concerns add to this challenge.

(2) **Complexity of Financial Markets**:
- **Non-Linear Relationships**: Financial markets are influenced by a complex web of interrelated factors. The relationships between these factors are often non-linear and can change over time.
- **Market Efficiency**: Efficient Market Hypothesis suggests that current asset prices reflect all available information. Predicting future movements based on historical data can be challenging as past patterns might not predict future movements.
- **Regime Shifts**: Economic conditions and policies can change, leading to regime shifts. Models trained on data from one economic regime might perform poorly in another.
- **Black Swan Events**: Unpredictable events (like the 2008 financial crisis or the COVID-19 pandemic) can dramatically shift economic trends, rendering existing models ineffective.
- **Behavioral Factors**: Economic and financial decisions are often influenced by human behavior, which can be irrational and hard to predict. Modeling these behaviors accurately is a significant challenge.
- **Non-stationarity**: non-stationarity refers to data whose statistical properties, (e.g., mean and variance), change over time, making it challenging to model and predict due to evolving trends, cycles, and unexpected events like market crashes or economic booms. **Non-Gaussianity**: non-Gaussianity of random variable distributions reflects irregular, often extreme events, like market crashes, which are not well-described by the normal distribution's bell curve, leading to heavier tails and more pronounced risks than Gaussian models would suggest.

(3) **Regulatory and Ethical Considerations**:
- Financial models are subject to regulatory scrutiny. Machine learning models, often seen as 'black boxes', can raise concerns about transparency, accountability, and fairness.

- Ethical concerns, such as the risk of amplifying biases in financial decision-making, are also significant.

(4) **Time Series Analysis Challenges**: Economic and financial data are typically time-series, which have their own challenges like autocorrelation, trend/cycle extraction, and handling of non-stationary data.
(5) **Generalization and Scalability**: Models trained on data from specific markets or periods may not generalize well across different contexts or times.
(6) **Feedback Loops**: Actions based on model predictions can influence the market, creating a feedback loop that can invalidate the model's assumptions.
(7) **Model Interpretability**: There's often a trade-off between model complexity and interpretability. In finance and economics, understanding the 'why' behind a prediction can be as important as the prediction itself.

3. Large Language Models

4. Knowledge Graphs

Knowledge graphs (KGs) store structured knowledge as a collection of triples

$$KG = \{(h, r, t) \subset \epsilon \times R \times E\}$$

where - E is a set of entities - R is a set of relations

Domain-specific KGs are construtued to represent knowledge in a specific domain (e.g., medical, biology, and finance)

5. Merging large Language models and knowledge graphs

5.1. **Pros and cons of LLMs.**

Large language models, pre-trained on large-scale corpora, have shown great performance in many natural language processing (NLP) tasks. By increasing model size, LLMs with billions of parameters (e.g., ChatGPT and PaLM2) have shown a surprising emergent ability and generalizability.

Despite their success in many applications, LLMs have been criticized for several limitations.

Advantages of LLMs

- General knowledge
- Language processing
- Generalizability

Limitations of LLMs

- Black-box: lack of interpretability
- Implicit knowledge: LLMs are black-box models, which fall short of capturing and accessing factual knowledge
- Hallucination
- Indecisiveness (reasoning happens through probabilistic process)
- Lack of domain-specific knowledge
- Lack of new knowledge

5.2. **Pros and cons of KGs.**

A potential solution is to incorporate knowledge graphs (KGs) into LLMs.

Some advantages of KGs are:

- Structural knowledge
- Accuracy
- Decisiveness
- Interpretability

- Domain-specific knowledge
- Evolving knowledge

Limitations of KGs include:

- Incomplete knowledge
- Lack language understanding
- Unseen facts

LLMs and KGs are two inherently complementary techniques, which should be unified into a general framework to mutually enhance each other.

6. Prediction model

6.1. **World model.**

6.1.1. *Agent.*

6.1.2. *Partially observable environment.* We assume that agent's sensors are not able to give information about the entire environment, but only about part of it. This can be due to noisy or inaccurate sensors, or to some parts of the state are missing from the sensor data

6.1.3. *Belief state.* Since the world is partially but not fully observable we need to maintain an internal state keeping track of the world. We need to model the world states (e.g., by enumeration or with formulas) and maintain a belief state on how likely each possible state of the world is, and we use probability theory to quantify the degree of belief.

6.1.4. *Action.* We assume that our system doesn't have ways to act on the world, but only it needs to estimate the state of the world and predict the next ones.

6.1.5. *Time slices.* We view the evolution of the world as a series of time slices (i.e., "snapshots"). For simplicity, we assume that the length of the intervals are the same. Each time slice contains a set of random variables: some random variables are not observable (unknown, hidden), e.g., the state of the world $\underline{\boldsymbol{X}}_t$, and other random variables are observable, and are called evidence $\underline{\boldsymbol{E}}_t$

Note that uncertainty over continuous time can be modeled by stochastic differential equation (SDEs). Discrete time models are discrete approximations to SDEs and so the discreteness assumption is not limiting.

6.1.6. *Handling of time.* TODO Time is handled making each quantity function of time.

We assume that the interval between slices is fixed, so we can label times with integers starting from $t = 0$. Evidence starts arriving at $t = 1$ $\underline{\boldsymbol{X}}_{[a:b]}$ represents a set of variables in $[a, b]$ - E.g., $U_{[1:3]}$ corresponds to the variables U_1, U_2, U_3

6.1.7. *Transition model.* TODO Use Belief state + transition model Predict how the world might evolve in the next step

The transition model specifies the probability distribution of the next state of the world $\underline{\boldsymbol{X}}_t$, given all the previous values:

$$\Pr(\underline{\boldsymbol{X}}_t | \underline{\boldsymbol{X}}_{0:t-1})$$

6.1.8. *Markov assumption for transition model.* In general the current state $\underline{\boldsymbol{X}}_t$ depends from a growing number of past states:

$$\Pr(\underline{\boldsymbol{X}}_t | history) = \Pr(\underline{\boldsymbol{X}}_t | \underline{\boldsymbol{X}}_0, \underline{\boldsymbol{X}}_1, ..., \underline{\boldsymbol{X}}_{t-1})$$

The Markov assumption is that the current state depends only on a finite fixed number of previous k states:

$$\Pr(\underline{\boldsymbol{X}}_t | history) = \Pr(\underline{\boldsymbol{X}}_t | \underline{\boldsymbol{X}}_{t-1:t-k})$$

For instance in a first-order Markov process the current state depends only on the previous state, and not on any earlier states

$$\Pr(\underline{\boldsymbol{X}}_t | history) = \Pr(\underline{\boldsymbol{X}}_t | \underline{\boldsymbol{X}}_{t-1})$$

so that a state provides enough information to make the future conditionally independent of the past. Sometimes the Markov assumption is exactly true (e.g., in the case of a random walk), other times it is a good approximation, depending on the domain.

6.1.9. *Time invariance for transition model.* In general the transition model depends on the value of t, requiring to specify a different distribution for each time step. We assume that the changes in the state of the world are time-homogeneous, i.e., the laws that govern the world don't change over time E.g., this assumption combined with a first-order Markov process yields

$$\Pr(\underline{\boldsymbol{X}}_t | history) = \Pr(\underline{\boldsymbol{X}}_t | \underline{\boldsymbol{X}}_{t-1}) = f(X_{t-1}) \forall t$$

6.1.10. *Sensor model.* In general the model to measure the evidence holds

$$\Pr(\underline{\boldsymbol{E}}_t | \underline{\boldsymbol{X}}_{[0:t]}, \underline{\boldsymbol{E}}_{[1:t-1]})$$

Under mild hypothesis, the model can be assumed to follow a form:

$$\Pr(\underline{\boldsymbol{E}}_t | \underline{\boldsymbol{X}}_t)$$

6.1.11. *Prior probability distribution at time 0.* The world model requires to specify the initial conditions $\Pr(\underline{\boldsymbol{X}}_0)$.

6.1.12. *Joint distribution.* Given a transition model following a first-order Markov process and sensor model, and the prior distribution we can infer the joint distribution overall all the variables:

$$\Pr(\underline{\boldsymbol{X}}_{[0:t]}, \underline{\boldsymbol{E}}_{[1:t]}) = \Pr(\underline{\boldsymbol{X}}_0) \prod_{i=1} \Pr(\underline{\boldsymbol{X}}_i | \underline{\boldsymbol{X}}_{i-1}) \Pr(\underline{\boldsymbol{E}}_i | \underline{\boldsymbol{X}}_{i-1})$$

6.1.13. *Update belief state.* TODO Use sensor model + percepts to update belief state

7. Description of the architecture

7.1. **Offline pass.**

- Step 1: generate domain-specific knowledge graph
  - The knowledge graph contains informations like:
    * Type of dependencies (e.g., causal, correlation)
    * Formula representing the type of relationship (e.g., sign, linear)
    * Strength of the relationship (in terms of belief)
  - We developed automatic ways (e.g., using LLMs) to generate a knowledge graphs of the domain
  - Human-in-the-loop approach to "regularize", inspect, and improve it

7.2. **Online pass.**

- Step 1: User poses a question to system
- Step 2: An LLM parses the question and extracts a relevant subset of the KG associated to the topic of the question
- Step 3: The subset of KG is converted into a BN that will answer the question
- Step 4: Relevant data (e.g., time series) needed to evaluate the BN is retrieved
- Step 5: The BN is evaluated, e.g., using KaizenFlow, in terms of values and uncertainties
- Step 6: The answer to the question is returned together with estimates about the error and the epistemological conviction

8. Building Bayesian model

9. Running the model

References

[1] Trace, *Decentralized Proving, Proof Markets, and ZK Infrastructure* (June 2023), available at `https://figmentcapital.medium.com/decentralized-proving-proof-markets-and-zk-infrastructure-f4cce2c58596`.