Trusting Artificial Intelligence (AI) for Healthcare

Synopsis

According to Health IT Analytics, "As demand for applications for AI in healthcare grows steadily as the technology rapidly advances, the potential for AI to enhance clinical decision support, chronic disease management and population health efforts has been checked by concerns over model bias and fairness" [i].

This paper addresses the application of causality [ii] and the use of causal AI in the modelling of healthcare applications, and how this branch of AI addresses the concerns of AI trustworthiness in the sector. A recent case study on using causal AI for healthcare claims leakage and fraud detection is included in the paper.

Causal AI is the only way to understand and successfully intervene in the world of cause and effect [iii]. There is a bias tendency to overestimate the strength and importance of the causal object and underestimate that of the effect object in bringing about the outcome [iv]. Causality fosters attribution, allowing us to attribute responsibility, credit, or liability.

Glossary

- **Attribution**—assigning responsibility, credit, or liability based on an event
- **Causal AI**—an AI model based on causality and not correlation, akin to human thought
- **Causality**—the relationship and reasoning between cause and effect
- **Causality inference (CI)—t**he process of identifying and understanding causal relationships between variables or events resulting from altering a causal model

- **Causal discovery (CD)—**the analysis and construction of models that depict the inherent cause and effect relationships within the data
- **Causal fairness—**the lack of bias in the resulting decision making from an outcome
- **Causal force—**cumulative directional effect of factors that affect time series trends
- **Causal graphs—**visualizes causal relationships between variables or events
- **Causal ML—**combination or endowment of causality with machine learning
- **Causal RL—**combination or endowment of causality with reinforcement learning
- **Computational linguistics—**training machines to understand questions and language
- **Confounder—**an influencing variable in causality that causes spurious correlations
- **Counterfactual Analysis—**an intervention to a causal model as to "what if" scenarios
- **Correlation—**dependence on a statistical relationship, whether causal or not
- **Deep learning (DL)—** subset of AI and machine learning for complex decision making
- **Ensemble modelling—**combining results of several analytical models into one spread
- **Experimental data—**the ability to control or intervene with variables in a model as an experiment to see if any causality is involved in the outcomes produced
- **Explainable AI (XAI)—** a process to allow humans to understand the results from AI
- **Feature Sets—**knowledge datasets produced by subject matter experts for causal AI use
- **Generative AI (GAI)—** AI capable of generating text, images, and videos from prompts
- **Graph knowledge databases—**a graphical representation of relationships of variables
- **Knowledge value chain—**process of capturing, storing, and reusing knowledge
- **Machine learning (ML)—** using data and algorithms to make machines learn like humans
- **Observational data—**the ability to observe certain variables to determine correlation
- **Reinforcement learning—**AI that trains automated agents to make real-time decisions
- **Variables—**an object, event. time series, or data item being measured by causal AI

Overview: Keeping the Human in the Loop

Healthcare and insurance are real-world applications that benefit from causality, where knowledge can be captured, stored, and reused by domain experts across healthcare ecosystems. Causality allows understanding of root causes of events, improving business decision making. By using human reasoning within machine learning, it explains event phenomena, identifying relationships between cause and effect and what influences the outcome.

Used by philosophers throughout the ages, the causality concept has now melded with the AI world and can be endowed to all branches of AI, generating all possible outcomes to mitigate and solve real-world issues.

By understanding causes, healthcare professionals learn from past experiences to avoid negative consequences and shape positive outcomes. Scientists systematically create, test, and challenge hypotheses to advance innovative frameworks, making causality integral to machine learning and statistics. This invokes probability theory, graphical models, and sensitivity analysis to assess the robustness of causal inferences to different assumptions affecting the causal relationship.

The downside of AI, such as trustworthiness, data quality, discrimination, bias, and hallucination (inaccurate results) are resolved by an intuitive discipline called causal fairness.

This enables causality to support decision makers beyond black box AI models and integrate more smoothly into organizations.

There is a distinction between causal inference and causality (Pearl) [v]. Causality is a philosophical concept of an event (the cause) being attributed for producing another event (the effect). Causal inference uses experimental and/or observational data and logical reasoning via computational analysis to quantify the strength of causal effects. It aims to determine causal relationships and effects between variables (events).
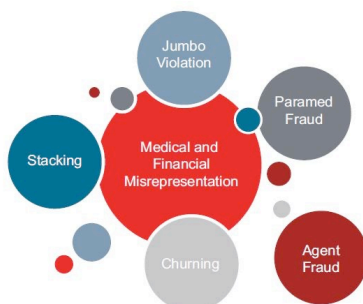
Healthcare Sector Overview
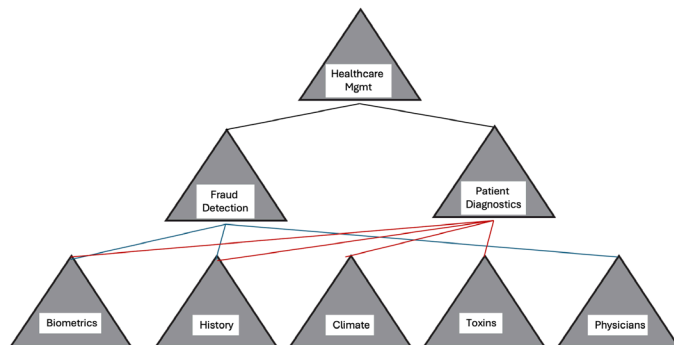
Source: Eumonics [vi]

The healthcare sector has much to gain from exponential technology and a confluence of AI and blockchain technology. Challenges exist in fraud and according to the International Organization for Medicare Anti-Fraud, health insurance fraud leads to a global annual loss of $260 billion, equivalent to 6 percent of total global health expenditure. [vii]

The global healthcare cyber security market size is expected to reach $56.3 billion by 2030 and expand at a CAGR of 18.5 percent over the forecast period [viii] due to increased cyber threats and data breaches caused by lack of data integrity. For the insurance industry, claim settlement and reconciliation times increase in line with fraud, and lack of medical data portability restricts both patient and insurer choices.

The following chart from RGA shows the magnitude of healthcare in insurance fraud [ix].

Source: RGA

In general, the stakeholders of the healthcare sector have existed in silos, making it difficult to address challenges. There is a necessity to move to an ecosystem approach as shown on the right-hand side of the diagram (courtesy of Guardtime Health). [x]
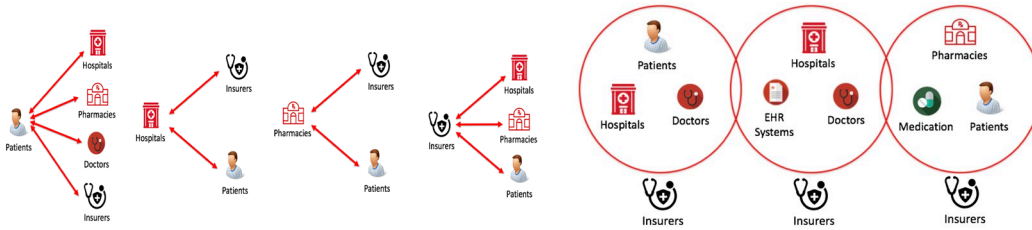
Source: Guardtime Health

Insurers can expand provider networks to a free-market economy, so hospitals compete on price, and there needs to be immutability of electronic healthcare records (EHR) to create trust in the network, reducing fraud.

This means transformation of healthcare systems using data analytics and smart devices based on improved patient outcomes, integrating multiple care providers across a broader healthcare ecosystem. This will centre around patient consent and control of personal health data using causal ML for large-scale medical data analysis.

The blockchain approach will electronically share prescriptions with pharmacies to mitigate against duplication or collusion between patient and doctor and encouraging outcome-based contracting/pricing. The pharma supply chain is then managed by accountability and visibility for improved clinical trials.

Meeting the challenge of population health and disease interception requires access to large volumes of data such as real-world data and evidence. This engages the patients beyond clinical trial completion where regulators require explanation of risk management and insurers' evidence of effectiveness and preserving privacy.

Patients and insurers expect a high quality of care from healthcare providers, leading to personal treatment plans. Following COVID-19, new products emerged, such as healthcare securitisation of pandemic risks (bonds) with granular real-time parametric data and surgery outcome insurance. Multi-disciplined datasets need to be added to the models such as air quality, as adding these variables can make the causal predictions more accurate.
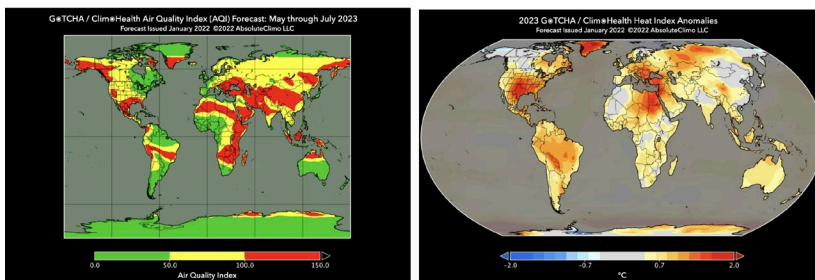
Source: AbsolutcClimo [xi]

Causality is crucial to healthcare, understanding disease cause, treatment effectiveness, clinical decision support, and precision medicine. Causal models can generalize across different populations so diverse groups receive tailored treatments. Precision medicine uses causal analysis to determine personalized healthcare approaches. Experimentation using causal models enables outcomes, allowing engineered solutions to be evaluated before

implementation when fully randomized experiments are impractical to conduct. In fraud detection, causality isolates causal factors for abnormalities rather than relying on correlations.

Moving From Correlation to Causality

Correlation tests for a relationship between two variables; but two variables moving together do not identify whether one variable causes the other to occur, hence the phrase "correlation does not imply causation" [xii]. It is important to understand the dynamic shift here as correlation is a statistical association that measures increasing or decreasing trends between two variables (events).

Causality is an empirical relation between two variables where a change in one (the cause) brings about a change in the other (the effect). Causation in AI is preferred to correlation as it explains why variables interact in a predictable and rational manner. Correlation analysis is statistical, being managed through high-performance computing power. Causal analysis adds rational explanation connecting the variables to real-world scenarios that define the risk.

Stochastic and actuarial modelling in the past has used copulas [xiii], not causality, to determine correlation in random variables, such as correlating an earthquake risk with a stock market downturn. Similarly, AI/ML rely on statistical correlations to infer causation.

Humans understand causation through dynamic intuition on how the world works and the ability to explain it. This identifies the root cause of an event by understanding the effects of any variables that triggered it, providing a deeper explanation of their true relationship.

Causal inference determines both the independent effect of an event and draws causal conclusions from data, such as A happened before B, B did not happen, C, D, E…Z and any other variable could not have caused B, providing a more human-like analysis. Causal relationships are directional as the cause precedes the effect but correlations can go either way and can exist without a cause-effect relationship.

Causal AI Methodology

Organizations that do not capture knowledge have loss of capitalized value. Causal AI is based on computational linguistics which is the computer modelling of natural languages whether written, verbal or visual. This is not translating languages, but capturing formal structures and rules to acquire the reasoning to train, interpret, understand and ask human questions. Computational linguistics capture, store, and reuse knowledge (intellectual capital), especially causal knowledge which understands the relationship between cause and effect and is developed with hypothesis-testing methodologies which are aligned with scientific methods.

This then formulates a hypothesis based on direction and magnitude vectors such as, "When Magnitude Vector > X, the price is likely to decline", or "When Magnitude Vector < X, the price is likely to increase". Computational linguistics are more compelling than methods based on statistical methods (linear regression), which includes large language models and deep learning/neural nets.
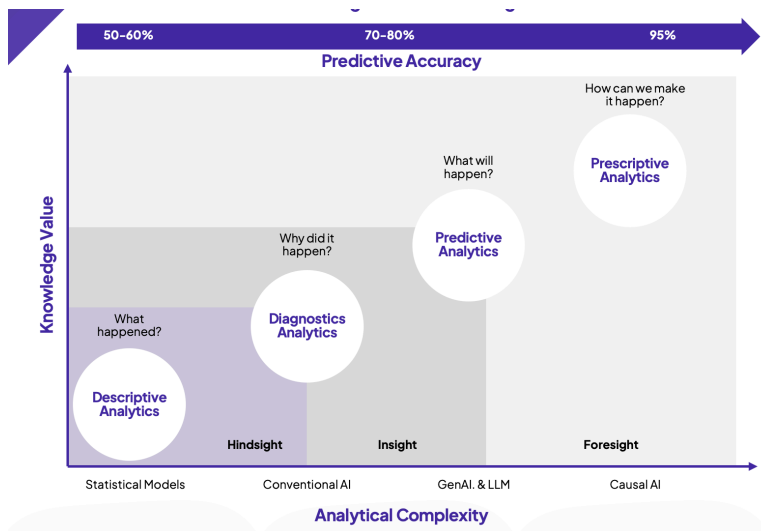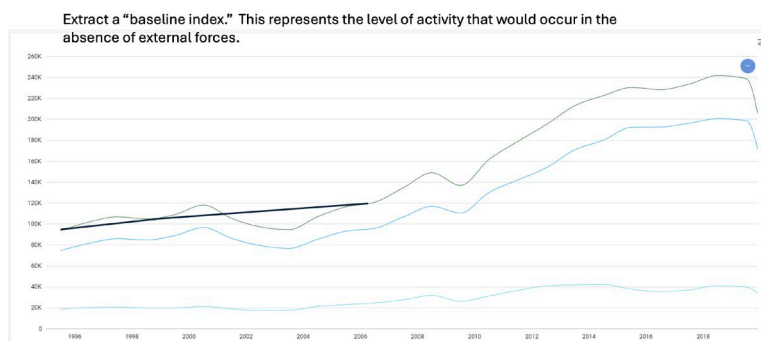
A hypothesis test explains a causal relationship based on data samples and determines if the hypothesis *explains* historical observations or *predicts* experimental outcome. If it does neither, the hypothesis is rejected.

Domain expertise can create knowledge data samples, known as feature sets. Human augmentation selects the best outcomes to improve the forecasting and integrated business planning by informing users how the decision by the model can be altered through minimal changes to the input features. This defines new content of causal factors such as indexes, baseline forces, benchmarks, and visualizations, all of which can be used in enterprise solutions.

Knowledge that has been captured for the domain is associated with the actual real-world behaviour being modelled and then users can continuously refine and improve the knowledge to make it as close to the real world as possible. This can then be used to explain what has (is) happening, predict what will happen next based on domain intent, and optimize responses to undesirable predictions and shape future outcomes.

For the healthcare case study, an index was created for each of the causal risk factors identified to use as a measure. Once created, a baseline index represents the mean in the absence of causal forces. The sequence is shown below, courtesy of Eumonics.
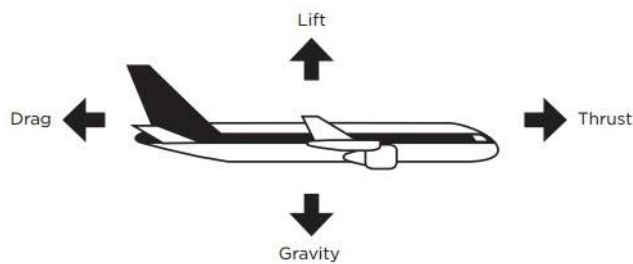

Extract a "baseline index." This represents the level of activity that would occur in the absence of external forces.

Baseline forces are important as they are the digital assets that become the enterprise intangible assets based on knowledge and can be used in parametric insurance.

Identify causes of inflection points in the baseline. Define a taxonomy of baseline forces.
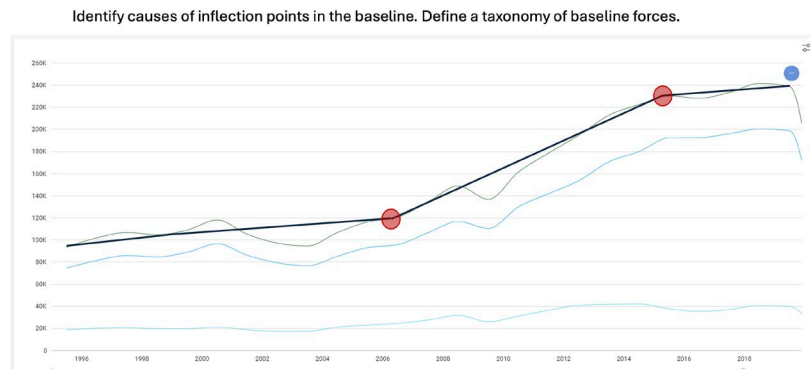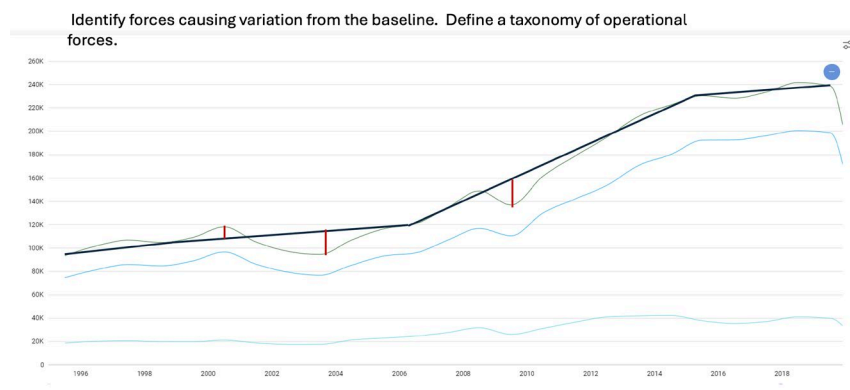
Source: Eumonics

The variations from the norm are then identified and known as operational forces. These can be upward or downward forces reflecting events; inflation, for example, would be a downward force, but interest rate drops an upward force, based on observational data. Like the forces on an aeroplane, if you quantify all the forces, you can then predict the position of the plane at any point of time.
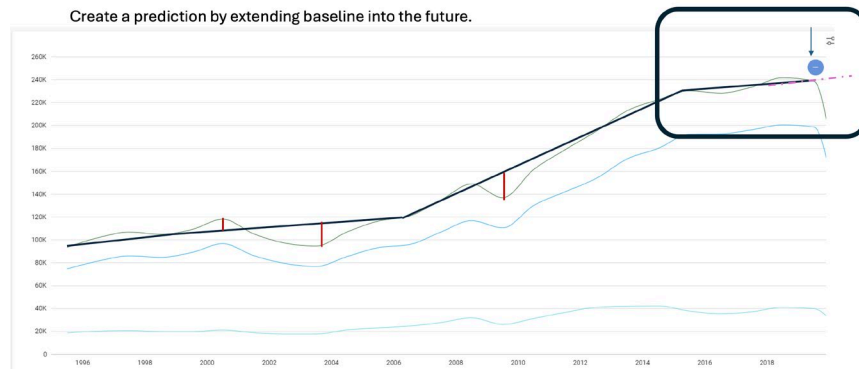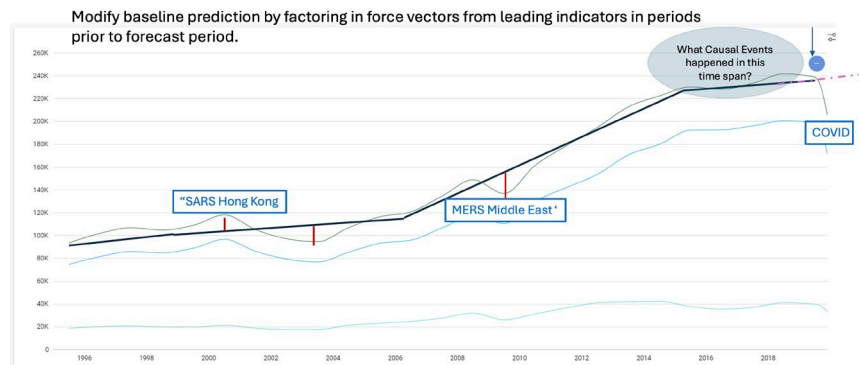

Source: Wikipedia

Identify forces causing variation from the baseline. Define a taxonomy of operational forces.

Source: Eumonics

Predictions are then created by extending the baselines into the future, as seen by the pandemic example.

Create a prediction by extending baseline into the future.

Source: Eumonics

Modify baseline prediction by factoring in force vectors from leading indicators in periods prior to forecast period.

Source: Eumonics

The causality approach is unique and differs from other kinds of AI. Statistical correlation is not the way humans think, but that is how AI has looked at solving problems with no explanation. Life is a multiple of variables and causality uses human reasoning. Statistics were used in the past as there was no causal technology mature enough, so identifying two variables together and linking them with statistical algorithms to interpret up and down forces was accepted, even though causality itself has been practiced since Plato [xiv].

Causal AI has been around since the 1950's and emerged from space research [xv] and was used in moon landings. Statistical analysis and experimentation could not be done so outcomes had to be generated from domain experts and then the best outcomes selected and then shaped to fit he mission before implementation. Criteria and evaluation metrics (risk scores) are related to the extracted results. If statistical methods put a spacecraft into orbit, thousands of experiments would be required in transporting, landing the spacecraft , departing, transporting back to earth, and tens of thousands of experiments to land the spacecraft safely on earth.

Multiple metrics in a holistic approach are needed to quantify causal understanding. Input datasets should include observational and interventional/experimental data and causality evaluation has to be based on multiple relevant datasets to test generalization. Benchmarking the transportability of causal conclusions to new domains and populations is crucial to understand their applicability beyond the training data.

Healthcare Causal Case Study

The following case study was presented to the Society of Actuaries in Asia in May 2024.[xvi] (courtesy of Vulcain.AI [xvii].)

It deals with experimental proof of concept related to uncovering leakage, or fraud, within a historical claims database using a causal AI platform, allowing rapid and accurate data harmonization, modeling, and hypothesis testing. Back testing was done to validate the model to the relevant predictive accuracy.

During assessment of the claims data, baseline values were established for each medical diagnostic code, connecting agents, hospitals, and doctors in one dataset as a claims path which, when analyzed, finds anomalies not detected with simple claim analyses. These anomalies are then represented as risk scores.

As more subject matter experts and data are added, additional possible magnitude of fraud vectors can be discovered by understanding the role that hospitals, agents, and doctors play in the leakage process. Claims data was harmonized, ingested, and analysed over several iterations. Explanations for anomalies were around characteristics for patients (age, gender, demographics), doctors, hospitals, treatment protocols, and exogeneous factors around climate, air quality, and environmental toxicity.

Using causal AI, three directional indexes were created for use in downstream models accessible through a dashboard, plus observations about potential patient treatment plans and pharmaceutical effectiveness.

The following indexes were created:

- Claims suspicious index, which risk scores the legitimacy of a claim
- Agent suspicious index, which assesses agent activity
- Ancillary cost index, providing insights on overall healthcare costs

Claims Suspicious Index

This will make claims teams more focussed on suspicious claims as claim patterns are being monitored in real time, applying predictive models to near- and long-term claims fraud.
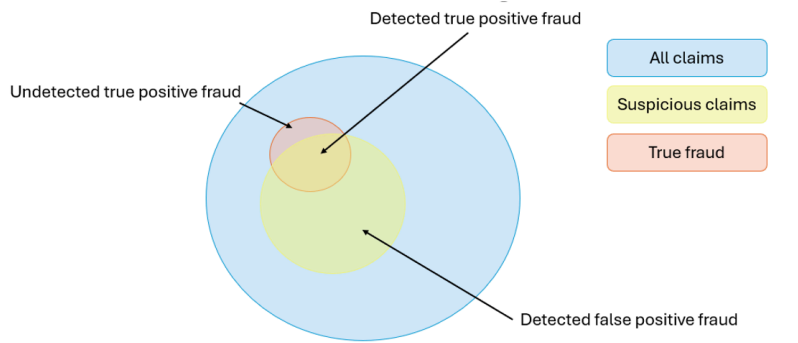
The red section shows the true leakage provided in the data or missed true fraud claims. The index was deliberately kept broad initially to capture as much suspiciousness as possible. The objective is to identify as many of the suspicious claims as valid. It is necessary to shows how an individual claim compares to others with the same diagnostic code.

The metric value of the claim here is recorded to the right as the blue line. The baseline for this measure is the black line, showing the mean, and each of the blue dots represent other claims with that diagnostic code.

This claim is a statistical outlier compared to the mean, giving a high suspiciousness index. Using this data in conjunction with hospital/doctor data, risk scores can be developed.
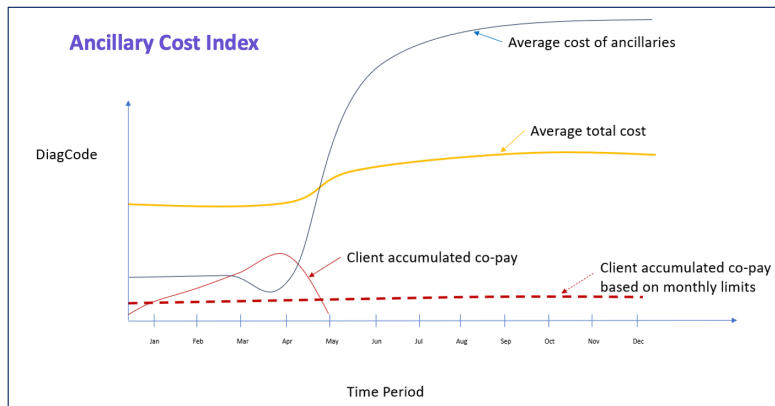
Agent Suspicious Index

This identifies the agents that cause the most leakage by monitoring agent patterns and behaviours in real time, applying predictive models to near- and long-term claims fraud.

Ancillary Cost Index

Source: Vulcain.AI

This demonstrates how average ancillary costs could fluctuate over time using time series causal analysis. Ancillary services support patient primary medical services, including diagnostic testing, laboratory services, imaging studies, rehabilitation therapy, medical equipment, and supplies.

This index provides insights into the financial impact of ancillary services on overall healthcare costs, identifies trends in ancillary service utilization and spending, and informs strategies for cost containment and optimization of healthcare resources.

As can be seen, it identifies that once the client's co-pay is reached, you get an increase in cost. This leads to a potential decision that could spread the co-pay throughout the year, providing a benchmark of overall expense costs assisting in the pricing mechanism for insurance products.

Conclusions From the Case Study

This leakage and fraud detection project should be expanded to include a larger dataset of claims to increase accuracy and further validate the index and associated causal models. Dedicated subject matter leaders, including the actuarial data science team, will help to clarify project goals, priorities, future dashboard implementation, insights, recommendations, process optimizations, and change management. Additional indexes should be hypothesized and tested for the use cases on patient and doctor effectiveness, hospital cost anomalies, and duplicate submissions and payments.

A patient treatment plan effectiveness strategy can be used to assess the success and efficiency of treatment plans implemented by healthcare providers for individual patients. This provides insights into the efficacy of treatment interventions, patient outcomes, adherence to clinical guidelines, and overall quality of care delivered. Treatment plan evaluations require large and diverse patient population datasets to predict accuracy, outcomes, and relevance.

A pharmaceutical effectiveness strategy can be used to assess the efficacy and safety of pharmaceutical drugs. The comprehensive metric factors in various factors, including the drug's therapeutic benefits, side effects, dosage regimen, patient compliance, and overall impact on health outcomes and clinical trials.

Better decisions can be made on identifying cost savings, eligibility of physicians and hospitals, capping claims, optimizing treatment efficiency, and forecasting future trends around procedures, costs, and outcomes. The next step would be to integrate the causal process with the underwriting for real time pricing.

Why Causal AI in the Healthcare Study?

The choice at this stage of AI development was obvious and shown in the chart created by the author. Over time, as causality is endowed on generative AI, the two disciplines will grow closer together; but at the time of this paper, generative AI (GAI) could not do this case study.

Causal AI	Generative AI
Assumptions explicit in causal models	Assumptions implicit in algorithms/data
Effects generalize beyond training data	Relies on identically distributed test data
Causal graphs encapsulate invariance	ML predictions can change unpredictably
Methods quantify and mitigate unfair bias	Can amplify bias in the training data
Estimates remain valid with interventions	Difficult to transport beyond training
Data-driven learning with human logic	Learn and generate data without causality
Infer causality through experimentation	Relies on statistics and correlations
Counterfactual analysis and "what if"	No counterfactual capability
Real-world understanding	No real-world understanding
Ability to apply attribution	Operates solely on correlations
Human hints limit the search	Complex search optimising no causality
Verticalized and subjective	Neither verticalized nor subjective
Based on computational analysis	Based on large language modelling

Importance of Determining Causality in Time Series

Time series analysis benefits from causal models for the examination of dynamic processes unfolding over time and tests the models' ability to discern short-term versus long-term effects.

Data in the time series are temporally related and when statistical analyses are performed on tabular data, it is assumed that the observations are independent, and expectation of future events does not depend on past ones [xviii]. However, analysing on the raw data measurements of a time series can lead to wrong conclusions [xix].

Take the following commonly used example in the causal graph below. Air pollution can increase in the summer and respiratory deaths increase in winter, so is time of year the root cause of air pollution and mortality?
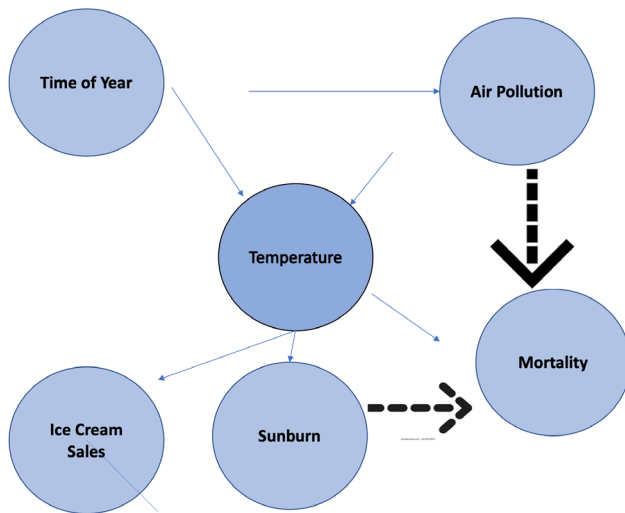
Looking at the problem on a time series year by year and aggregating mortality and pollution across different populations shows higher pollution causes higher mortality, exposing that the time of year has confounded the correlation.

In causal inference, a confounder is a variable that can negatively influence the model and cause spurious associations as a confounding variable C is causally associated with both the exposure (X) and outcome (Y) but C is not on the causal pathway of X and Y.

ML often makes an inconsistent decision as it cannot tell a spurious correlation from a true causal effect. Looking at this example, it can infer that a rise in temperature correlates with sunburn and ice cream sales, but ML can give a spurious correlation between ice cream and sunburn. When scaling to the size of the internet, the volume of these spurious correlations is problematic and a human hint is needed to limit the search to avoid exabytes of false data.

To accurately predict the root cause of pollution on mortality, causal inference and discovery are required to derive the answer to manage time of year confounding variable and misleading correlations plus adding richer datasets. Having managed the confounders, a model now exists that accurately looks at pollution and deaths along a time series, and the model becomes a sandbox to freely ask questions not possible with statistical correlation.

Causal graphs capture interactions and relationships between variables and map causal pathways to outcomes of interest. A causal graph can show that some mortality is caused directly by air pollution and others by extreme temperatures and can be constructed by domain experts using observational data, or by a combination of human knowledge and ML algorithms.

Intervention gives the ability to make counterfactual predictions about what would happen if certain variables were changed to alter the causes of a particular outcome but holding fixed other relevant variables or confounders. They can be applied to both observational and

experimental data and are a tool to understand cause-and-effect relationships in complex systems under controlled conditions.

Counterfactual questions, such as what will happen if air pollution levels rise or drop by certain percentages, will aid better business decisions in the climate change arena to reduce mortality. In the case study, the ancillary index was created using the time series causal analysis.

Counterfactual Analysis: Challenging the Causal Model

Counterfactual questions are the basis of philosophy challenging a hypothesis such as "what would the world look like if the INTERNET had not been invented". This is the integration of empirical observation and experience (knowledge) laid down by philosophers down the ages, giving rise to explicit and not implicit predictions. Philosophy links causation to explanation. There is a need to have more overlap between philosophy and computer (and data) science.

A counterfactual explanation describes the smallest change in a prediction that alters that prediction to a predefined output, thus shaping the event. Causal AI evaluates these explanations and shows how and why the best ones were selected so regulators get access to the different outcomes to show fairness and an unbiased approach—for example, to pricing.

Counterfactual theories of causation can be explained in terms of, "If A had not occurred, C would not have occurred". Every business sector benefits from a better understanding of risk; and when a significant event occurs ("black swan"), questions are asked about how the loss might have been averted or mitigated.

One question that should be asked is how the loss might have been worse. This can be applied to a cyber-attack on a healthcare facility and may shape mitigation of a future, more serious attack, as the bad actors will be thinking how to make it worse. This is known as a downward counterfactual.

This is described very well in Lloyds of London/RMS Reimaging History report [xx], which specifically applies downward counterfactuals to catastrophes. This 2017 report actually argues the case for causality in the insurance sector but preceded causal AI developments. The report highlights that in statistical analysis, historical data is fixed rather than one possible version of many that could have occurred if various influencing factors had been different.

Causality also mitigates inbuilt bias inherent in risk models that are based on a single standard dataset used by all modellers when exploring long tail risks by facilitating complex explanations and deeper understanding of future risks.

Risk analysis is an observational experience, but laboratory experiments in the past have only been possible on a small scale. This changes with causal AI where the maximum knowledge can be gained from historical experience and domain experts. As the report states, "This includes not just what happened, but what almost happened and what might have happened". Downward counterfactual risk analysis provides a useful tool for regulators to stress-test, check, and back-test risk models. Statistical correlations do not support counterfactual reasoning.

Causality Trustworthiness

Causal models must be trusted for responsible deployment and continued reliance on the model. This addresses legal, ethical, and practical concerns regarding AI safety and transparency and trust intensifies when models are extracted without human intervention.

Causality models by definition hold trustworthiness characteristics that can be affirmed by human reasoning. Data quality plays a critical role, as potential biases and errors/omissions in observational data can undermine the accuracy of causal inferences.

Identifying the causal structure and assumptions from the given data is crucial as this can impact the validity of the conclusions. Confounding must also be addressed to ensure the outcome variables are properly measured and accounted for to avoid biased estimates.

Digitization has enabled efficient handling of data by powerful processors, combined with the use of advanced query languages and ML models. Causal inference must operate on a large volume of data that is beyond human processing capabilities and needs to accommodate velocity, volume, value, variety, and veracity to conduct large-scale experiments and uncover cause-effect relationships within real-world scenarios.

Large observational datasets have facilitated exploration of complex causal structures, with thousands of variables, previously too costly to analyse. Passive observational studies utilizing extensive patient datasets have emerged as viable alternatives to randomized experiments when identifying treatment effects, thanks to the availability of adequate sample sizes.

The combination of multiple datasets, encompassing diverse sources like electronic health records, social media, and mobile data, can provide complementary insights; but this increased capability raises privacy risks. Applying data integrity while preserving privacy not only mitigates against cyber-attacks but also allows provenance, non-repudiation, attribution, and immutability to the underlying data in the causal models.

The following diagram shows how a holistic data integrity program can be initiated using blockchain hashkey cryptography.

Source: Guardtime [xxi]
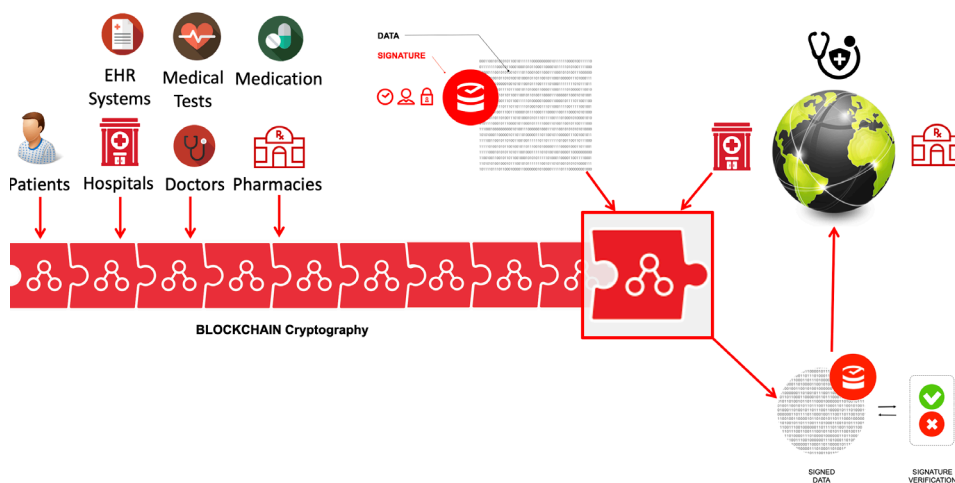
Healthcare data must overcome lack of interoperability and, to achieve trust, having a single version of the truth with undisputed provenance and veracity. Healthcare data can be analysed with permission without physically sharing the data which, if needed, can also be transferred without compromise in the case of payments and billing.

Causality and Explainable AI (XAI)

XAI provides transparency and understandability to the decision-making process of complex AI systems, often perceived as black boxes. By combining causality and XAI, a comprehensive causal visual explanation of AI decision-making is achieved by identifying the causal relationships between variables that lead to these decisions. The XAI approach unlocks the black box mysteries of AI and allows for algorithmic explanation for regulatory use.

XAI can generate counterfactual explanations. Models embedding causality identify critical information in datasets, discarding irrelevant or outlier correlations and improving predictions leading to trustworthy AI solutions. Causality explanation characteristics increase the trustworthiness and good for decision making and allows for transportability of knowledge.

Implementing Causality on Graph Databases

Graph databases show relationships between models and provide a solid platform for causal analysis by tracing causal pathways and connectivity. By identifying interconnected causal networks, it becomes possible to determine all entities indirectly impacted by a specific event or condition.

Graph databases also enable the simulation of what-if scenarios, as they allow for the modification of causal link attributes. This facilitates the exploration of how a causal network may behave under varying conditions or interventions.

Graph databases also provide a unified structure that facilitates the integration of diverse causal data types, such as experimental, observational, temporal, and spatial data into a single repository, taking a holistic comprehensive approach to studying causality (see example below.) [xxii]
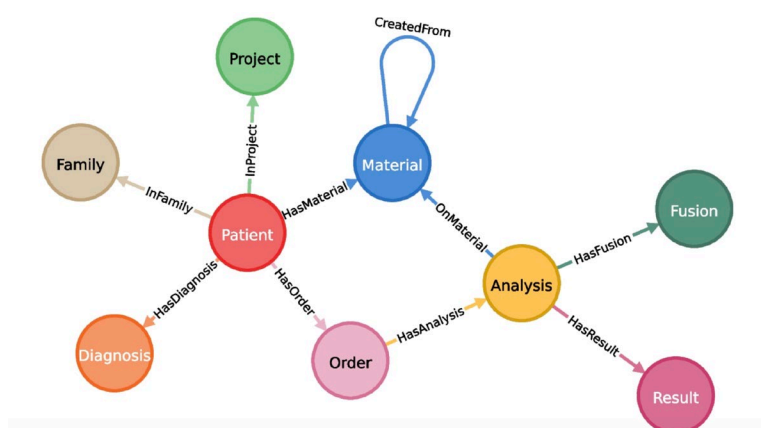
Source:Graph4Med [xxiii]

A potential cause must occur before an alleged effect, establishing a clear temporal order demonstrating a statistical association between the potential cause/effect variables in the data. A feasible mechanism based on scientific theory explains how cause leads to the effect.

The gold standard for establishing causality is experimental data, where the cause is intentionally manipulated (intervention) to prove its impact on the effect.

Observational data is more passive but valuable as domain data. The model can also be adjusted by a dose-response relationship, whereby the magnitude of the cause is adjusted to correspond to larger or smaller magnitudes of the effect.

Data quality is critical, necessitating accurate, complete, and multi-variate data with minimal missing information. Measurement error, selection bias, and attrition can undermine causal conclusions. Replication is needed, as causal conclusions carry more weight when multiple independent studies using different datasets and methods yield consistent effects, minimizing the likelihood of chance discovery.

Conclusions and Future Directions

Intelligence is the ability to acquire knowledge and skills, but AI is the ability of machines to acquire the same. The need is to apply human reasoning to AI to avoid bias and inaccurate results. The healthcare sector is in the crosshairs of AI adoption. AI is associated with challenges of trust such as bias and black box concerns. These challenges are holding back the positive developments so the ability to create trustworthiness for data and the resulting decision-making models is crucial. The solution for instilling this trust is causal AI, a branch of AI based on the philosophy of causation. This can be endowed on the other AI subfields such as ML, DL, RL, and GAI to bring transparency and rational explanations to the decision makers and apply human reasoning to the solutions ensuring the right balance of human and machine.

A gap still exists between human and AI/ML causality that needs to be narrowed. AI/ML relies on statistical correlations in data to infer causation, whereas humans understand dynamically through intuition and common-sense knowledge with the ability to consider uncertainties like confounding variables.

Most AI/ML approaches currently take a limited, static view of causation, so concerns over the quality of real-world data, ethics, and privacy dominate discussions for better governance and responsible AI. Total automation is not the answer; the human should remain in the loop, with data collected from subject domain experts who remain involved monitoring through the entire knowledge value chain.

Causality enables counterfactual thinking, where philosophy and data science meet with the tendency to consider the past and question what could have been if a different decision had been made.

This moves away from one fixed point in history based on one dataset to adopting a counterfactual perspective on new historical experience and exploring other possible realisations of history (Woo) [xxiv].

Casual AI is the only branch of AI that can do counterfactual analysis and visualise it through use of causal graphs. The case study in this paper highlights the causal process; but it is early days for this discipline, and Gartner predicts two to five years for it to reach full adoption and integration to other fields of AI, as shown in the 2023 hype cycle [xxv] which shows a great deal of attention being paid to generative AI.
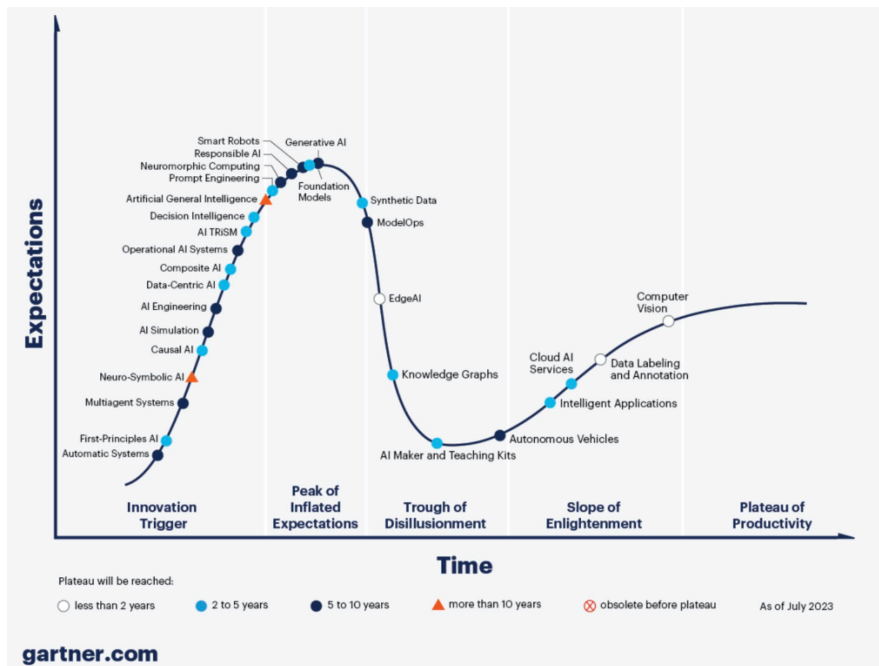
Valuable results were obtained from the case study, but more observational data input and subject matter domain expertise is required to increase the accuracy of predictions and the resulting baseline forces/indexes which will form the enterprise system assets to embed the causality. Experimentation is involved, much like quantum computing, where all the outcomes are generated quickly as in a laboratory environment and humans then select, intervene, and improve the best outcomes, allowing the machine to do all the heavy lifting.

As casual AI naturally holds all the trustworthiness characteristics of AI, it makes sense for the healthcare, insurance, and other business sectors to endowing causality to other branches of AI so they inherit those capabilities. In the opinion of the author, this leads to a bright future as a confluence of quantum computing and causal AI will bring ethics, privacy, and transparency to the results and address the healthcare sector requirements—for example, cancer cures and new drug innovation.

This is all dependent on having the right data integrity, cyber integrity, provenance, and quality in the observational data used in the causal process leveraging observed data patterns, so missing data can be imputed using ML algorithms and human assumptions.

That means embedding blockchain deep in the architectural plumbing as a real-time smoke alarm for data tampering and mitigation of cyberattacks, and a need for all organizations to set up responsible AI foundations aligned with current and future regulations.

Looking to the future, it behoves business to see through current AI hype and create a strategy to combine the best of human thinking and computational power across the enterprise. Deep learning (DL), for example, is a subscience of machine learning (ML) and ML algorithms can enable the attachment of causal models to data.

These models allow predictions about intervention effects on outcomes so estimating counterfactuals and hypothetical outcomes is achievable through ML algorithms. By utilizing data from control groups, this could estimate the outcome for individual patient treatment. By controlling confounding variables with causal time series analysis, genetic profiles, and medical histories, with consent, can be employed to predict the treatment's impact on a specific patient.

Genomics and clinical data identify causal links between genetic variants and diseases. This would require access to a considerable amount of datasets and use the ensemble model approach which is very common now in the ML space.

Ensemble methods combine multiple predictive models to improve overall performance, make more accurate predictions, and reduce uncertainty in causal inferences. This averages plausible estimates from diverse models encoding alternative assumptions. Identifying the most impactful causal relationships can be accomplished by analysing the magnitude of their effects. This leads to an evolution of data analytics from purely stochastic, to conventional AI , to predictive analytics (including generative AI) and to prescriptive analytics (causal AI) leading to causality over statistical analysis and correlation. This shows a shift from sole use of Monte Carlo simulation models and probability theory based on historical data, to future looking models based on ensemble modelling blending various machine learning models and accessing big data casual analysis models consisting of domain knowledge blocks.

Reinforcement learning (RL) is a branch of ML and a fast-growing discipline in daily lives today. These are automated agents and co-pilots that learn to make decisions by interacting with an environment, removing repetitive and menial tasks from the human, which involves the agent taking actions, receiving feedback in the form of rewards or penalties, and using this feedback to improve its decision making over time.

RL allows automated agents to learn optimal behaviour directly from experience to solve complex real-world challenges through trial-and-error interaction and real-time learning. This is a growing field due to emerging of usages of large language models (LLM). Causality and RL are connected through their shared aim of decision making through modelling systems and both address challenges by leveraging observational data. Causality RL addresses this challenge by integrating causal inference methods into the RL framework.

Agents use causal models to represent the causal relationships between actions and environmental changes, enabling them to infer the causal effects of their actions. Counterfactuals play a role in reasoning about the effects of alternative actions.

Causality RL offers improved decision-making and explanations. Causal reinforcement learning is another promising area, involving the integration of causal reasoning into automated agents such as co-pilots so agents better understand how their actions propagate consequences in the environment.

To conclude, it is not unusual for causal AI companies to have access to millions of datasets, some proprietary and others in the public domain. Data is the most important intangible asset and providing data quality and integrity today used in AI training will determine the usage in future generations.

It is important to get the foundational data correct today so that future generations are using the correct real-world data and the bias and hallucination challenges in AI today are ironed out to ensure we do not lose control of data and maintain that provenance for the future.

Humans should not be scared of AI and should embrace it as it will create new horizons and opportunities for all. As menial tasks are made redundant or automated, humans can use their full potential. Causality is the right approach to keep the human in the loop, explain and shape the outcomes, and take advantage of scientific progress. This is where the crossover of computer science and philosophy bring in powerful mechanisms for causal analysis.

References

The author references the work and publication of Robert C. Whitehair and Victor R. Lesser entitled, "A Framework for the Analysis of Sophisticated Control in Interpretation Systems" [xxvi] that enabled the case study in this paper.

[i] https://healthitanalytics.com/features/overcoming-struggles-to-define-algorithmic-fairness-in-healthcare

[ii] https://en.wikipedia.org/wiki/Causality#:~:text=Causality%20is%20an%20influence%20by,partly%20dependent%20on%20the%20cause.

[iii] https://ssir.org/articles/entry/the_case_for_causal_ai

[iv] https://pubmed.ncbi.nlm.nih.gov/16478304/#:~:text=This%20bias%20is%20termed%20the%20causal%20asymmetry.

[v] https://statmodeling.stat.columbia.edu/2014/01/13/judea-pearl-overview-causal-inference-general-thoughts-reexpression-existing-methods-considering-implicit-assumptions/

[vi] https://www.eumonics.com/about-us/

[vii] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11061682/

[viii] https://finance.yahoo.com/news/healthcare-cyber-security-market-analysis-133700332.html

[ix] https://www.rgare.com/knowledge-center/article/investigating-life-insurance-fraud-and-abuse

[x] https://guardtime.com/health

[xi] https://absoluteclimo.com/

[xii] https://www.jmp.com/en_hk/statistics-knowledge-portal/what-is-correlation/correlation-vs-causation.html#:~:text=Correlation%20tests%20for%20a%20relationship,correlation%20does%20not%20imply%20causation."

[xiii] https://en.wikipedia.org/wiki/Copula_(probability_theory)#:~:text=Copulas%20are%20used%20to%20describe,to%20grammatical%20copulas%20in%20linguistics.

[xiv] https://link.springer.com/chapter/10.1007/978-94-015-9747-0_1

[xv] https://ntrs.nasa.gov/citations/20230015678

[xvi] https://www.actuaries.org.hk/

[xvii] https://vulcain.ai/

[xviii] https://www.sigmacomputing.com/resources/learn/what-is-time-series-analysis

[xix] https://www.researchgate.net/publication/349236961_Causal_Inference_for_Time_series_Analysis_Problems_Methods_and_Evaluation

[xx] https://assets.lloyds.com/assets/reimagining-history-report/1/Reimagining-history-report.pdf

[xxi] https://guardtime.com/

[xxii] https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-022-05092-0

[xxiii] http://graph4med.cs.uni-frankfurt.de/

[xxiv] https://www.casact.org/sites/default/files/2021-07/Counterfactual-Disaster-Woo.pdf

[xxv] https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle

[xxvi] https://web.cs.umass.edu/publication/docs/1993/UM-CS-1993-053.pdf

7.2024

David Piesse
CRO of Cymar

About the Author:

David Piesse is CRO of Cymar. David has held numerous positions in a 40-year career including Global Insurance Lead for SUN Microsystems, Asia Pacific Chairman for Unirisx, United Nations Risk Management Consultant, Canadian government roles and staring career in Lloyds of London and associated market. David is an Asia Pacific specialist having lived in Asia 30 years with educational background at the British Computer Society and the Chartered Insurance Institute.