A Custom GPT Health Recommender System Informed by Causal Bayesian Networks and Authoritative Ontologies

Hengyi Hu, Larry Kerschberg

Submitted to: JMIR AI
on: June 15, 2024

Table of Contents

A Custom GPT Health Recommender System Informed by Causal Bayesian Networks and Authoritative Ontologies

Hengyi Hu[1] PhD; Larry Kerschberg[2] PhD

[1]George Mason University Fairfax US
[2]George Mason University` Fairfax US

Corresponding Author:
Hengyi Hu PhD
George Mason University
4400 University Dr
Fairfax
US

Abstract

Background: ChatGPT and other large language models (LLMs) are trained on extensive text data; they learn patterns and associations within the training text without an inherent understanding of underlying causal mechanisms. Establishing causation necessitates controlled experiments and observations, and as of May 2024, ChatGPT lacks access to experimental data and the capacity to learn analytical models from data. Recent advancements from OpenAI enable the creation of custom Generative Pre-trained Transformers (GPT) models using their GPT Builder. These custom GPTs can be tailored with causal knowledge from Causal Bayesian Networks (CBN), thereby producing a knowledgeable, health-recommender system with causal expertise. This system is not only easily accessible for patients, clinicians and other users, but also has the potential to significantly improve healthcare outcomes.

Objective: This paper presents a practical solution –a custom GPT model integrated with Causal Bayesian Networks (CBNs) informed by Authoritative Medical Ontologies (AMOs) as prior foundational knowledge. AMOs are robust biomedical ontologies that encapsulate the expert knowledge of their creators. By utilizing structured information contained within these ontologies, we can generate an informed CBN which can be used to improve a custom GPT as a health recommender system. These enhanced GPTs offer profound insights into cause-and-effect among co-morbid symptoms within the disease domain.

Methods: To demonstrate our recommender system, we learn a CBN using NIMH data for patients of Alzheimer's Disease and augment this with causal knowledge from the International Classification of Diseases Version 10 Clinical Modification (ICD-10-CM). We compute the CBN using the Max-Min Hill-Climbing (MMHC) algorithm. We generate two separate CBNs using MMHC to compare predictive accuracies between a baseline and a CBN modified with causal mechanisms from ICD-10-CM. Our previous research using this method has resulted in a modified CBN that reflects the causal claims in the AMOs and agrees with both the AMOs and the observational dataset. With this causal model, we build a custom GPT using OpenAI's GPT Builder.

Results: The custom GPT contains both potentially causal and correlations among symptoms, as well as conditional probabilities for these relationships. The GPT will also contain knowledge of causal mechanisms from ICD-10-CM, extended information regarding symptom variables, and references to existing literature regarding comorbid Alzheimer's Disease symptoms. Furthermore, because the modified CBN model establishes potentially causal relationships among symptoms which can be verified in existing epidemiological research, we can verify that the custom GPT also establishes these causal relationships. This creates a GPT that agrees with the modified CBN, which is a representation of existing subject matter expertise in the disease domain.

Conclusions: To obtain our CBN, we've used a previous methodology which obtains ordered variable pairs from authoritative ontology ICD-10-CM as prior expertise. In our past research, a CBN that is learned using MMHC can be improved significantly by considering prior sources of knowledge, if the algorithm is modified appropriately. This source of prior knowledge can be validated in existing literature as a sequence of events in AD progression, specific causal mechanisms among comorbid symptoms, and conditional probabilities from a Bayesian Network. The resulting modified network provides insight into the causal relationships expressed in the AD data and takes advantage of the expertise and knowledge contained in the AMOs. Since inferring causality from a CBN does not exist in a vacuum, the relationships within the CBN, regardless of the strength of the conditional probabilities, must be explored further. A custom GPT extends the causal knowledge within a CBN with general knowledge of symptoms and diseases, providing a tool capable of suggesting causal inference based on the analysis of real patient data. With a LLM, uncertainty is present but reasoned with, as it is present but reasoned with in the prior, posterior, and likelihood in Bayes Theorem. Moving forward, we would like to explore using ChatGPT to produce ordered variable pairs and to automate the validation of potentially causal information.

(JMIR Preprints 15/06/2024:63297)
DOI: https://doi.org/10.2196/preprints.63297

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

A Custom GPT Health Recommender System Informed by Causal Bayesian Networks and Authoritative Ontologies

Hengyi Hu
Department of Information Technology
George Mason University
Fairfax, VA
hhu2@gmu.edu

Larry Kerschberg
Department of Computer Science
George Mason University
Fairfax, VA
kersch@gmu.edu

Abstract— **ChatGPT and other large language models (LLMs) are trained on extensive text data; they learn patterns and associations within the training text without an inherent understanding of underlying causal mechanisms. Establishing causation necessitates controlled experiments and observations, and as of May 2024, ChatGPT lacks access to experimental data and the capacity to learn analytical models from data.**

Recent advancements from OpenAI enable the creation of custom Generative Pre-trained Transformers (GPT) models using their GPT Builder. These custom GPTs can be tailored with causal knowledge from Causal Bayesian Networks (CBN), thereby producing a knowledgeable, health-recommender system with causal expertise. This system is not only easily accessible for patients, clinicians and other users, but also has the potential to significantly improve healthcare outcomes.

This paper presents a practical solution –a custom GPT model integrated with Causal Bayesian Networks (CBNs) informed by Authoritative Medical Ontologies (AMOs) as prior foundational knowledge. AMOs are robust biomedical ontologies that encapsulate the expert knowledge of their creators. By utilizing structured information contained within these ontologies, we can generate an informed CBN which can be used to improve a custom GPT as a health recommender system. These enhanced GPTs offer profound insights into cause-and-effect among co-morbid symptoms within the disease domain.

To demonstrate our recommender system, we learn a CBN using NIMH data for patients of Alzheimer's Disease and augment this with causal knowledge from the International Classification of Diseases Version 10 Clinical Modification (ICD-10-CM). We compute the CBN using the Max-Min Hill-Climbing (MMHC) algorithm. We generate two separate CBNs using MMHC to compare predictive accuracies between a baseline and a CBN modified with causal mechanisms from ICD-10-CM. Our previous research using this method has resulted in a modified CBN that reflects the causal claims in the AMOs and agrees with both the AMOs and the observational dataset.

With this causal model, we build a custom GPT using OpenAI's GPT Builder. The custom GPT contains both potentially causal and correlations among symptoms, as well as conditional probabilities for these relationships. The GPT will also contain knowledge of causal mechanisms from ICD-10-CM, extended information regarding symptom variables, and references to existing literature regarding comorbid Alzheimer's Disease symptoms. Furthermore, because the modified CBN model establishes potentially causal relationships among symptoms which can be verified in existing epidemiological research, we can verify that the custom GPT also establishes these causal relationships. This creates a GPT that agrees with the modified CBN, which is a representation of existing subject matter expertise in the disease domain.

Keywords—Patient data, data mining, data management, Bayesian networks, causal inference, causal networks, causality, healthcare data, healthcare information technology, health recommender system, ontology, Alzheimer's Disease

I. Introduction

A. Custom Generative Pre-trained Transformers and Causal Limitations of Large Language Models

A custom Generative Pre-trained Transformers (GPT) are a type of Large Language Model (LLM) specifically tailored to cater to a specialized function [1]. Unlike a standard ChatGPT model that is trained on a broad range of internet text to handle a wide variety of topics, custom GPTs are fine-tuned with targeted datasets. This fine-tuning process involves training the pre-trained model further on a specific set of text that is reflective of additional expertise. This expertise could be legal language, technical manuals, or even advanced analytical models such as a Bayesian Network and conditional probabilities. This specialization allows the custom GPT to excel in specific domains or applications, thereby providing more accurate and relevant outputs when dealing with specialized content such as causality and causal mechanisms [2].

GPTs and other LLMs are great tools for natural language processing, but they have well-known limitations in establishing causality, thereby impacting their effectiveness [3], [4], [5]. This is rooted in how Artificial Intelligence (AI) and Machine Learning (ML) deal with pattern recognition in text rather than explicit causal mechanisms and direct cause-and-effect relationships. A GPT does not understand causation because its responses are based on correlations in text. It makes causal statements based on correlations in its training data, but it has no inherent understanding of causation. For example, ChatGPT might say “smoking causes heart attacks” because smoking is a well-established risk factor for heart attacks. However, it does not inherently understand the confounding that occurs within the domain of vascular diseases. In this case, heart attacks are confounded by a multiple exogenous variables such as high cholesterol (atherosclerosis), high blood pressure, hypertension, cardiovascular disease, diabetes, and smoking – all of which are comorbid and causes of heart attacks which are well-documented [6], [7], [8].

Furthermore, ChatGPT can interpret and perform math calculations by converting text-based questions into mathematical operations [9] but cannot learn from any new experimental data without a computational API extension such as Wolfram [10]. This means it cannot learn a Bayesian Network with conditional probabilities if we give it new data. It also means that it cannot provide textual answers to new, qualitative mathematical problems. For example:

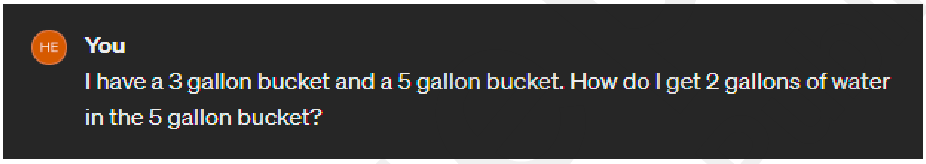
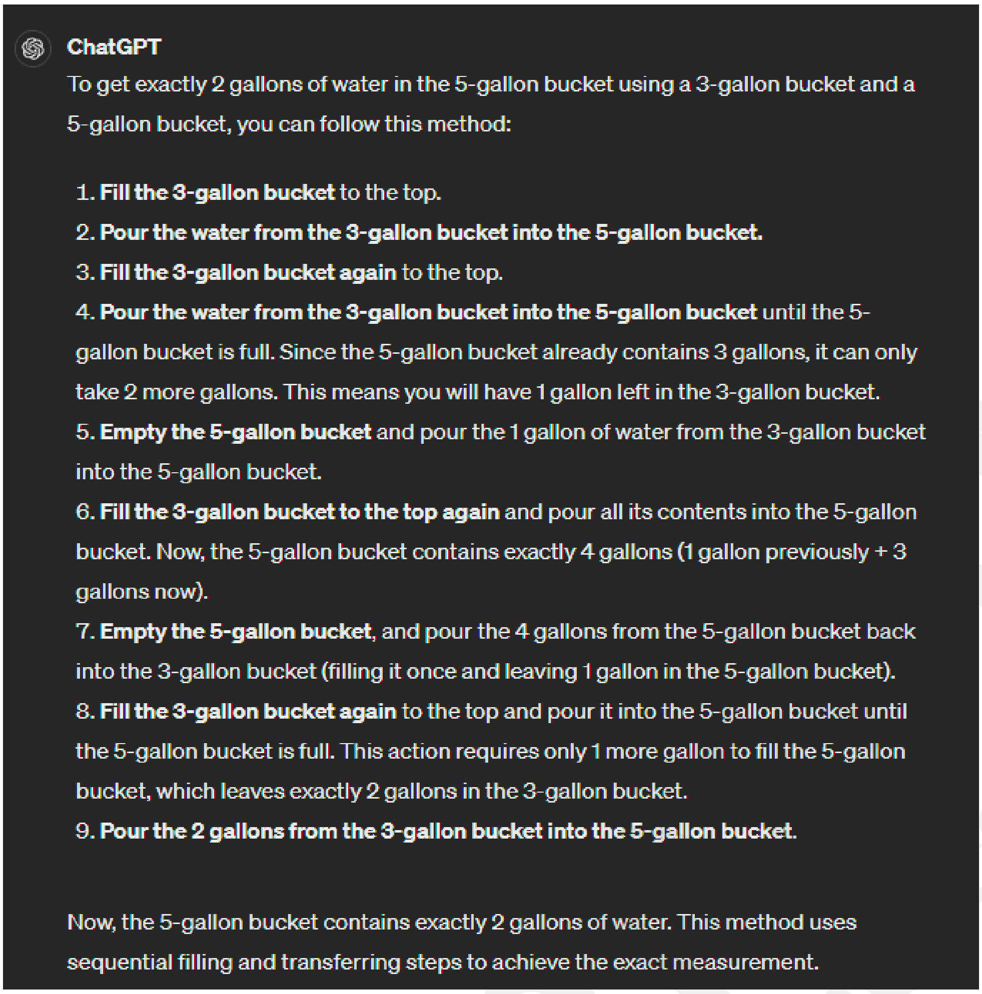


Fig. 1a: Simple Causal Query to GPT-4O

The correct answer requires understanding a simple cause-and-effect problem where the 5-gallon bucket is filled first. Since a LLM does not inherently understand causation, we receive a very long and convoluted answer from ChatGPT-4O based on text correlation from its training set:



ChatGPT

To get exactly 2 gallons of water in the 5-gallon bucket using a 3-gallon bucket and a 5-gallon bucket, you can follow this method:

1. **Fill the 3-gallon bucket** to the top.
2. **Pour the water from the 3-gallon bucket into the 5-gallon bucket.**
3. **Fill the 3-gallon bucket again** to the top.
4. **Pour the water from the 3-gallon bucket into the 5-gallon bucket** until the 5-gallon bucket is full. Since the 5-gallon bucket already contains 3 gallons, it can only take 2 more gallons. This means you will have 1 gallon left in the 3-gallon bucket.
5. **Empty the 5-gallon bucket** and pour the 1 gallon of water from the 3-gallon bucket into the 5-gallon bucket.
6. **Fill the 3-gallon bucket to the top again** and pour all its contents into the 5-gallon bucket. Now, the 5-gallon bucket contains exactly 4 gallons (1 gallon previously + 3 gallons now).
7. **Empty the 5-gallon bucket**, and pour the 4 gallons from the 5-gallon bucket back into the 3-gallon bucket (filling it once and leaving 1 gallon in the 5-gallon bucket).
8. **Fill the 3-gallon bucket again** to the top and pour it into the 5-gallon bucket until the 5-gallon bucket is full. This action requires only 1 more gallon to fill the 5-gallon bucket, which leaves exactly 2 gallons in the 3-gallon bucket.
9. **Pour the 2 gallons from the 3-gallon bucket into the 5-gallon bucket.**

Now, the 5-gallon bucket contains exactly 2 gallons of water. This method uses sequential filling and transferring steps to achieve the exact measurement.

Fig. 2b: Convoluted response to Simple Causal Query in GPT-4O

GPTs generate responses based on patterns and associations from its large dataset and lacks an understanding of its contents. Without a general understanding of cause-and-effect, it may generate plausible but factually incorrect or nonsensical responses called hallucinations [11]. Since GPTs do not inherently grasp causality, responses may sound correct given the context of the question but are not logically consistent with any understanding of a sequence of events.

Finally, the performance and reliability of GPTs are heavily dependent on the quantity, quality, and diversity of the training data [10]. Poor or limited training data can lead to gaps in knowledge, reduced accuracy, and a failure to generate appropriate responses in less common or unexpected scenarios. While ChatGPT-4O has a very large text training set, it will not contain any new insights into areas where we lack data or understanding of the data. GPT may present information in a different way based on correlations in the LLM, but not provide new insights or any new understanding of that information [12].

B. Causal Inference and Causal Bayesian Networks

Determining causality for diseases, comorbidities, and treatments requires an understanding beyond interpreting correlation and association, which is at the center of the data vectorization process in ChatGPT and LLMs [13]. While statistics gives us insight into the collected data, they do not address the underlying causes of the diseases, or the underlying causes of the adequacy of treatment. Statistics alone cannot measure: 1) causes and effects, and 2) how or why causes influence their effects [14]. Traditional statistical methods do not address the causal mechanisms behind why a treatment is successful, what caused the population to be sick, or if one symptom causes another as the disease progresses. Statistics alone are insufficient for establishing causality [15], [16].

There are three criteria that are generally viewed as necessary for identifying a causal relationship: 1) association (or correlation) between the variables, 2) proper sequencing based on a time order (or ordering of variables), and 3) the absence of confounding (non-spuriousness) of the associations [17], [18], [19], [20]. Establishing a causal relationship is further strengthened by the *identification* of a causal mechanism and the *context* for the relationship. A gold standard for inferring causality between two events is to determine which occurred first [21], [22], [23]. This is important for data modeling where there are internal (endogenous) variables to the model and external (exogenous) variables [14].

A classic understanding of correlation and regression is that association (or correlation) does not imply causation [24]. Having a correlation between two variables does not mean that a change in one variable causes change in another – as defined in Pearl's definitions of causality. While establishing correlation is straightforward with classical methods such regression analysis, establishing cause-and-effect is much more difficult. Having a strong correlation, however, establishes a relationship between a dependent and independent variable and provides an opportunity to investigate causality. For example, a dataset may show a correlation between smoking and drinking, but we don't know if smoking causes drinking or visa-versa. This relationship will need to be further examined to determine if a cause-and-effect relationship can be established.

In 1998, Judea Pearl described how causal relationships can be inferred from cross-sectional data if certain assumptions are made regarding the underlying process of data generation [25]. If the data variables are observed in a linear sequence, we can assume that a variable *X* preceding another within that sequence is a causal indicator of the second variable *Y*. The causal indicator *X* in this case, has the potential of being a cause of *Y* or assigning a value to *Y*. In our research, we adopt Judea Pearl's definition of causation [14]:

1. *Variable X is a cause of variable Y, if Y in any way relies on X for its value.*
2. *Variable X is a direct cause of variable Y if X appears in a function that assigns values to Y.*

For example, the question "which symptoms are indicators for dementia for an Alzheimer's patient?" requires a causal model exploring causal relationships [26]. To explore causal relationships, we need: 1) a method to articulate causal assumptions, 2) a way to link structures of causal models to data, and 3) a method to draw conclusions based on causal assumptions in a model and in the data. The desire to infer causal relationships from structures such as a directed acyclic graph (DAG) has led to the creation of different inference models. Causal graphs generated from data can capture the probabilistic and causal properties of multivariate distributions [27]. These graphs capture a joint probability distribution using a graphical representation, which is used to visualize conditional dependence and independence [28].

A DAG capable of capturing probabilistic relationships among random variables is a Bayesian Network (BN). BNs are composed of a set of random variables $X=\{X_i, i=1,\dots,n\}$ and a DAG, denoted $G=(V,A)$ where *V* is the set of nodes and *A* is the set of arcs [29], [30]. Each *node* in *V* represents one *variable* in the data, and they are referred to interchangeably. The directed arcs in *A* that connect nodes in *V* are denoted as "$\rightarrow$" and represent direct dependencies between nodes. If there is no arc connecting two nodes, the corresponding variables are either marginally independent or conditionally independent, given a subset of the rest of the variables. As a result, each local distribution depends only on a single node X_i and on its parents, and $P(X)=\prod_{i=1}^{n}$ ¿¿ .

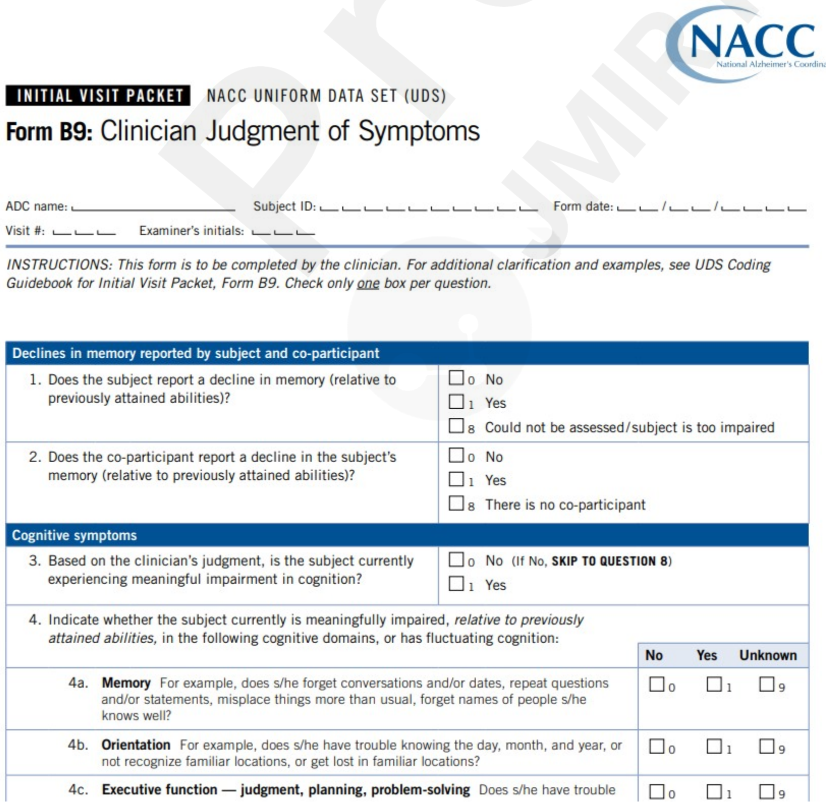
From a Bayesian perspective, given a Bayesian Network *B = (G,* X_i*)* and a dataset *D*, we can compute the posterior probability $P(B|D)=P(G,\theta|D$¿ $\propto$ $P(G|D$¿$*P(\theta|G,D$¿, where $P(G|D$¿ is the learning of the BN structure, and $P(\theta|G,D$¿ is the learning of the BN parameters. Fitting a BN to the data (learning) consists of two main steps [31]: 1) Structure learning: a single DAG structure is discovered which best fits the data according to a specific algorithm used, and 2) Parameter learning: determining the probability distributions of arcs *A* among nodes *V* in the DAG.

To determine the structure of a network based on data, conditional independence tests are used. Conditional independence is a key concept in Bayesian Networks [25] due to factorizations of joint probability distributions. Given a variable set *V*, two random variables *X*, *Y* are conditionally independent given *Z*, or $(X \perp Y \vee Z)$ if $\forall x, y, z : P(X=x \vee Y=y, Z=z) = P(X=x \vee Z=z)$; provided that $\forall z : P(Z=z) > 0$. Since conditional independence is a concept based on traditional statistical independence, if two variables *X*, *Y* are independent, then the joint distribution is the product of the marginals: $P(X=x, Y=y) = P(X=x) P(Y=y)$. If the two variables *X*, *Y* are dependent given conditioning on *Z*, then: $P(X=x, Y=y \vee Z=z) = P(X=x \vee Z=z) P(Y=y \vee Z=z)$

When a BN is used to infer causality and combined with known causal probability distributions between pairs of nodes, it produces a Causal Bayesian Network (CBN). CBNs are extensions of BNs in which potentially causal relationships are represented as conditional probabilities between nodes [32], [33]. For example, a directed edge $X \rightarrow Y$ indicates a single causal relationship between a pair of variables, from variable *X* to variable *Y*. Causal models which are used to infer causality are limited by the interpretation of causal mechanisms external to the data. A directed edge $X \rightarrow Y$ may denote a correlation (BN) or a causal relationship (CBN) with prior knowledge or additional verification.

C. National Alzheimer's Coordinating Center's (NACC) Uniform Data Set (UDS)

This experiment uses the National Alzheimer's Coordinating Center's (NACC) Uniform Data Set (UDS). The NACC UDS contains longitudinal data collected since September 2005 during standardized annual evaluations conducted at the National Institute on Aging (NIA)-funded Alzheimer's Disease Research Centers (ADRCs) across the country. Participants in the study represent the entire Clinical Core enrollment of the ADRCs, with cognitive status ranging from demented to mild cognitive impairment to cognitively normal.



NACC
National Alzheimer's Coordina

INITIAL VISIT PACKET NACC UNIFORM DATA SET (UDS)

Form B9: Clinician Judgment of Symptoms

ADC name: ______ Subject ID: ______ Form date: __/__/____

Visit #: ___ Examiner's initials: ___

INSTRUCTIONS: This form is to be completed by the clinician. For additional clarification and examples, see UDS Coding Guidebook for Initial Visit Packet, Form B9. Check only one box per question.

Declines in memory reported by subject and co-participant	
1. Does the subject report a decline in memory (relative to previously attained abilities)?	☐0 No ☐1 Yes ☐8 Could not be assessed/subject is too impaired
2. Does the co-participant report a decline in the subject's memory (relative to previously attained abilities)?	☐0 No ☐1 Yes ☐8 There is no co-participant
Cognitive symptoms	
3. Based on the clinician's judgment, is the subject currently experiencing meaningful impairment in cognition?	☐0 No (If No, **SKIP TO QUESTION 8**) ☐1 Yes

4. Indicate whether the subject currently is meaningfully impaired, *relative to previously attained abilities*, in the following cognitive domains, or has fluctuating cognition:

	No	Yes	Unknown
4a. **Memory** For example, does s/he forget conversations and/or dates, repeat questions and/or statements, misplace things more than usual, forget names of people s/he knows well?	☐0	☐1	☐9
4b. **Orientation** For example, does s/he have trouble knowing the day, month, and year, or not recognize familiar locations, or get lost in familiar locations?	☐0	☐1	☐9
4c. **Executive function — judgment, planning, problem-solving** Does s/he have trouble	☐0	☐1	☐9

Fig. 3: Form B9 Clinician Judgment of Symptoms for NACC UDS Dataset

The raw dataset provided by NACC has been pre-stratified to include only Alzheimer's Disease (AD) as the primary cause of their visit. The dataset spans 15 years and is collected using 18 individual forms[1], consisting of both self-reported historical patient data and etiologic diagnosis from clinicians. The raw data provided by the NACC is pre-stratified to include only variables where AD is the primary cause of the doctor's visit. This dataset contains over 74,000 patient entries and 720+ variables. From the raw data, a subset was selected which focuses on the clinician's assessment of the patients of AD. The following filters are applied:

- Selection of any records in the timeframe of September 2005 (the beginning of the UDS) to November 2020.
- Only variables obtained from Form B9: Clinician Judgment of Symptoms, and Form D1: Clinician Diagnosis are used. Form B9 generates a variety of variables focusing on cognitive, behavioral, and motor symptoms diagnosed by a clinician. Form D1 provides us with a diagnosis of dementia, which will be the target variable to test model accuracy.
- Only variables which are directly determined by a clinician in Forms B9 and D1 are used. Variables which are self-reported by the subject/patient are not used.
- Only complete records are used and any records which are coded "9 = Unknown" or "-4 = Not available" are excluded
- Non-binary variables, such as text descriptions, are excluded.
- Only variables collected using the original UDS questions are used. "NACC derived variables", or variables which are derived based on other UDS questions, are excluded.
- Variables for "Other" changes in cognition and behavior, where the specific symptom is not identified and the variable has a single value for "No" and "Unknown", are excluded.
- Summary variables, e.g., Question 3 in Form B9, "…is the subject currently experiencing meaningful impairment in cognition?" are excluded because the specific symptoms diagnosed by the clinician are listed in Questions 4a. – 4g.

Post-stratification, 26 variables and 4422 patient records remain. Each binarized variable for the symptoms pertains to the clinician's assessment of a patient of AD. Age is the only non-binary variable. The variable names and related description pertaining to Form B9 and D1 are listed in Table 1.

Table 1: Selected NACC UDS Variables for Patients of Alzheimer's Disease

Variable Name	NACC UDS Description
SEX	Gender, 1 = Male, 0 = Female
AGE	Age of patient at diagnosis, the difference between two other variables: BIRTHYR (year of patient's birth) and VISITYR (year of patient's visit).
DEMENTED	Subjects who met the criteria for all-cause dementia have a positive diagnosis. Subjects with normal cognition who did not meet the criteria for dementia have a negative diagnosis
COGMEM	Indicate whether the subject currently is meaningfully impaired, relative to previously attained abilities, in memory
COGORI	Indicate whether the subject currently is meaningfully impaired, relative to previously attained abilities, in orientation
COGJUDG	Indicate whether the subject currently is meaningfully impaired, relative to previously attained abilities, in executive function - judgment, planning, or problem-solving
COGLANG	Indicate whether the subject currently is meaningfully impaired, relative to previously attained abilities, in language
COGVIS	Indicate whether the subject currently is meaningfully impaired, relative to previously attained abilities, in visuospatial function
COGATTN	Indicate whether the subject currently is meaningfully impaired, relative to previously attained abilities, in attention or concentration
BEAPATHY	Subject currently manifests meaningful change in behavior - Apathy, withdrawal
BEDEP	Subject currently manifests meaningful change in behavior - Depressed mood
BEVHALL	Subject currently manifests meaningful change in behavior - Psychosis - Visual hallucinations

[1] https://naccdata.org/data-collection/forms-documentation/uds-3

BEAHALL	Subject currently manifests meaningful change in behavior - Psychosis - Auditory hallucinations
BEDEL	Subject currently manifests meaningful change in behavior - Psychosis - Abnormal, false, or delusional beliefs
BEDISIN	Subject currently manifests meaningful change in behavior - Disinhibition
BEIRRIT	Subject currently manifests meaningful change in behavior - Irritability
BEAGIT	Subject currently manifests meaningful change in behavior - Agitation
BEPERCH	Subject currently manifests meaningful change in behavior - Personality change
BEREM	Subject currently manifests meaningful change in behavior - REM sleep behavior disorder
BEANX	Subject currently manifests meaningful change in behavior - Anxiety
MOGAIT	Indicate whether the subject currently has meaningful changes in motor function - Gait disorder
MOFALLS	Indicate whether the subject currently has meaningful changes in motor function - Falls
MOTREM	Indicate whether the subject currently has meaningful changes in motor function - Tremor
MOSLOW	Indicate whether the subject currently has meaningful changes in motor function - Slowness
MOMOPARK	Were changes in motor function suggestive of Parkinsonism?
MOMOALS	Were changes in motor function suggestive of amyotrophic lateral sclerosis?

D. Algorithms and Algorithm Selection for learning Causal Bayesian Networks (CBN)

The first step in generating a BN is structure learning, which consists of an algorithm that generates a DAG which best represents the conditional independencies present in the data. This has been discussed in the literature using constraint-based, score-based, and hybrid algorithms [29], [34]. The second step is parameter learning. Parameters of the BN are calculated based on available data and by assigning joint densities to the edges between nodes of the DAG. A single edge is a conditional density based on parameters that are estimated using Bayesian estimation techniques (minimize loss or maximize utility in the posterior) and regularized maximum likelihood [33].

Since the 1980s, three approaches have emerged as viable methods for causal learning: 1) constraint-based, 2) score-based, and 3) hybrid. Constraint-based approaches search the data for conditional independence between variables and attempt to find a DAG that captures the corresponding d-separations [35]. In score-based methods, Bayesian approaches attempt to find a DAG which maximizes the likelihood of the posterior given a prior dataset [36], [37]. A third hybrid approach combines both approaches.

These algorithms share common assumptions, including: 1) a correspondence between nodes V and random variables X (two nodes cannot correspond to a single variable), 2) arcs between variables represent conditional dependencies, 3) all possible combinations of X are valid events, and 4) observations from X are independent and identically distributed (i.i.d), because a sequence or collection of random variables is i.i.d [38]. We now explore several established approaches that mitigate the difficulty of learning a CBN.

Constraint-based algorithms are based on the seminal work of Pearl on causal graphical models and his Inductive Causation (IC) algorithm [39], which provides a framework for learning the DAG of a BN using conditional independence tests under the assumption that graphical separation and probabilistic independence imply each other (the faithfulness assumption). All constraint-based structure learning algorithms share a common three-phase approach inherited from the IC algorithm. This can be summarized as: 1) optional learning of Markov Blankets, 2) learning neighboring nodes (e.g. parents and children), and 3) network construction and learning arc directions [35].

Score-based learning algorithms uses heuristic optimization techniques. A candidate DAG is generated then scored based on how well they fit the data. A typical method for constructing a DAG is to acquire Bayesian posteriors from a set of priors. In this method, the posterior probability of the DAG given the dataset is calculated using Bayes' rule where the score of the DAG given the data:

$$p(DAG \vee Data) = \frac{p(Data \vee DAG) * p(DAG)}{p(Data)}$$

Examples of score-based learning algorithms include: efficient caching using decomposable scores [40], parallel meta-heuristics [41], and integer programming [42]. Hybrid algorithms use both conditional independence tests and network scores. The former to reduce the space of candidate DAGs, and the latter to identify the optimal DAG among them. Some examples are Parent-Children (PC) [43], Grow-Shrink (GS) [29], [44], Incremental Association (IAMB) [45], Max-Min Hill-Climbing (MMHC) [46], [47], and Max-Min Parents & Children (MMPC) [47].

We will determine an algorithm to use to learn the BN by performing a k-fold (k=5) cross-validation for a set of popular BN learning algorithms. Using the `bn.cv` function in *bnlearn* and the NACC data selected from the previous section, we will use Log-Likelihood Loss [48] (`logl`): also known as negative entropy or negentropy, it is the negated expected log-likelihood of the test set for the BN fitted from the training set. A lower log-loss value means better predictions.

Table 2: Log-Likelihood Loss Scores for NACC dataset

Algorithm	Logl Score
Grow-Shrink	15.752
Hill-Climbing	14.612
Tabu	14.623
Max-Min Hill-Climb	14.493
Restricted Maximization	14.564
Hybrid HPC	14.468
Incremental Association	14.642

For the following experiment, the Min-Max Hill Climbing (MMHC) algorithm has been chosen due to its popularity as a hybrid algorithm for learning Bayesian Networks [46], [47]. MMHC is a popular hybrid (score and constraint-based) algorithm. It uses very little memory and can find models for very large datasets and state spaces. MMHC is a hybrid structural learning algorithm which utilizes the Max-Min Parents Children (MMPC) to restrict search space, and regular Hill-Climb (HC) to find the optimal network structure.

E. Orienting BN arcs for Causal Inference

The algorithms discussed in the previous section determine the structure of the network, the direction of the arcs within the structure, and the conditional probabilities of arcs between nodes. When a BN is used to infer causality and contains causal knowledge, it becomes a CBN where the arc direction between two nodes can specify a causal relationship where one node is the cause and the other is the effect. One of the unique challenges of orienting arcs within a CBN is obtaining a prior sequence or ordering among the symptom variables. There are several methods of reliably obtaining prior knowledge to assist in orienting arcs within a CBN. They can be categorized into three groups:

1. <u>Time-based, longitudinal sequencing of variables</u>. This is the most intuitive method where the diagnosis of each symptom variable has a recorded time period, time stamp, or some other time-based sequence [49]. This method is very common in epidemiology where the progression of a disease happens in a sequence over time [50]. Recent research has shown that the fit and accuracy of score-based CBNs improve noticeably with prior time-based sequencing information [33], [50], [51], [52]. This is done by obtaining a prior time-based sequence of symptoms from the data, then altering the algorithm used to generate the CBN based on the temporal sequence.
2. <u>Logical ordering of variables</u>. This method requires logical analysis and ordering of cross-sectional data variables where time-based sequencing is not available. Ordering can be intuitive, e.g. symptom progression based on age [53], or it can be logical. For example, symptoms variables precede treatments, but other demographic variables such as race, ethnicity, gender,

etc., precede symptoms.

3. Other data-based methods. These methods establish temporal sequencing or logical ordering based on the data, and includes collider testing [54], human perception of causal strength [55], and Error Reduction [56]. A patient's age can also be used determine the sequence of events [57].

Once a prior orientation of potentially causal relationships is established, it can be used to significantly improve conditional probabilities within a causal network learned from data by up to 20% in certain studies [51], depending on the accuracy of the prior orientation and the available data.

F. Authoritative Medical Ontologies (AMOs): ICD-10-CM

An ontology is commonly defined as a "specification of a conceptualization" in the context of knowledge and data sharing [14]. Information stored in modern ontologies are repositories for specific application domains, such as healthcare. AMOs contain lexicons used throughout the healthcare industry for patient diagnoses, medical research, regulation policies, and product development. Examples of AMOs include the Gene Ontology, NCI Thesaurus, SNOMED CT, ICD-10-CM, and MedDRA. These ontologies provide a method of standardizing domain knowledge in a variety of different but related biomedical fields. This paper utilizes causal information from ICD-10-CM[2], a system used by physicians and other healthcare providers to classify and code all diagnoses, symptoms and procedures recorded in conjunction with hospital care in the United States. ICD-10-CM is based on the International Classification of Diseases 10th revision (ICD-10) published by the World Health Organization. It uses unique alphanumeric codes to categorize diseases and symptoms. Clinicians, information technology experts, and other healthcare professionals in the U.S. use ICD-10-CM to store and retrieve diagnostic information of symptoms, diseases, and treatments. The current ICD-10-CM repository contains over 71,000 classes as of September 2019.

This paper utilizes our prior research of obtaining ordered variable pairs from ontological subsumption hierarchies and indexed terminologies in one or more Authoritative Medical Ontologies (AMOs). Ordered variable pairs from an AMO contain prior knowledge which can then be used to orient the arcs/edges in a CBN. Orienting arcs occurs after the baseline network has been established. This is a data-driven methodology, which contrasts with existing methodologies to create CBNs directly from ontologies [58], [59], [60] instead of from data.

G. Arc Agreement and Measuring Predictive Accuracy of Bayesian Networks

For the quantitative assessment, we will compare the CBNs based on the following:

1. Agreement of arcs between models: Use *compare()* function in the *bnlearn* package in *R* [61] to measure agreement between Modified and Baseline CBNs. This function counts the number of directed arcs that are the same (or different) between two networks.
2. Predictive accuracy: We compute the cross-validated Area Under the ROC Curves (AUC) of the Baseline and Modified CBNs. AUCs are used to summarize the Receiver Operating Characteristics (ROC) curve, which checks a model's predictive performance. The ROC is a probability curve, and the area under it represents a measure of separability. It tells us how much a model is capable of distinguishing between classes. At higher AUCs values, the model is better at predicting negative diagnoses (0s) as negative diagnoses, and positive diagnoses (1s) as positive diagnoses.

The *compare()* function utilizes a "target" network and "current" network. The "target"

[2] https://www.cdc.gov/nchs/icd/icd10cm.htm

network is taken to be "true" or as the "golden standard" network, and the "current" network will be compared to it. Three metrics are returned:

1. *True positive* (tp) arcs appear both in target and in current
2. *False positive* (fp) arcs appear in current but not in target
3. *False negative* (fn) arcs appear in the target but not in current

In order to visualize the performance of the CBNs, we can use the ROC (Receiver Operating Characteristics) curve and the AUC (Area Under the Curve) [62]. The ROC curve is a performance measurement for the classification problems at various threshold settings. ROC is a probability curve and AUC represents the degree or measure of separability. This metric measures how well the model is capable of distinguishing between classes. For binary datasets of patient diagnoses, 0s mean a negative diagnosis class and 1s mean a positive diagnosis class. The higher the AUC, the better the model is at predicting 0s as 0s and 1s as 1s.

FPR tells us what proportion of the negative class was incorrectly classified by the classifier. A higher TNR and a lower FPR is desirable since we want to correctly classify the negative class. In a ROC curve, a higher X-axis value indicates a higher number of False positives than True negatives. While a higher Y-axis value indicates a higher number of True positives than False negatives. The choice of the threshold depends on the ability to balance between False positives and False negatives. If the plot of the ROC curve (and the AUC) is higher in one model vs. another, then we can say that the model with the higher AUC score did a better job of classifying the positive class in the dataset.

H. Prior work in Causal Modeling, Methodology, and Generalizability

In our research, we have captured patient symptoms in an ontology module [63], stored associations (non-causal) among symptoms as relationships in a modular ontology [64], created a CBN to demonstrate its compatibility and translation to an ontology [65], and proposed improvements to medical ontologies by analyzing causation in patient data [66]. We have also established a methodology that utilizes an AMO as prior causal knowledge for learning a CBN [67].

BNs and ontologies have intrinsic compatibilities which enable them to be modeled after each other. For example, we can create a BN using the semantic information found in ontologies [68], [69], [70]. Even though a viable CBN can be derived directly from an ontology, this method relies on the accuracy of the ontology rather than an accurate representation of patient data. Deriving a BN from an ontology is possible because semantic representation of knowledge can be translated into components of a BN:

1. Nodes are represented as ontology classes;
2. Edges and structure are represented as relationships between classes, and;
3. Probability distributions are derived from data instances.

Conversely, ontologies can also be created from the structure of BNs [71]. This is achieved by extending the standard OWL ontology language to express conditional, probabilistic relationships between classes. Three extensions currently exist: *BayesOWL* [71][3], *PROWL* [72][4] and *OntoBayes* [73]. *BayesOWL* is a framework which extends OWL capabilities for modeling and reasoning with uncertainty. *OntoBayes* improves upon BayesOWL by supporting random variables with multiple values. Finally, *PROWL* further extends OWL where probabilistic concepts can co-exist with regular, non-probabilistic concepts. In each of these extensions, a set of rules is applied to transform the class hierarchy defined in an OWL ontology into a Bayesian network.

[3] http://semanticweb.org/wiki/Bayes_OWL.html
[4] http://www.pr-owl.org/

Methods for learning BNs directly from ontologies have also been proposed [33], [53], [58], [74], [75]. These methods discover causal structures depicted in ontologies, and then test the accuracy of the structure using observational data. To obtain a CBN directly from an ontology, the taxonomical structure is examined and then used to infer causality. This creates a CBN model based on ontology classes, relationships, and instances. The conditional probabilities in the resulting model are based on the number of instances, class properties, or the number of definitions in other ontologies. This method creates a CBN model that is based on pre-defined ontologies, rather than learned from data. Since they are based on a pre-defined ontology and not learned from data, it is not possible to measure them for conditional probabilities among symptoms or for predictive accuracy.

There are inherent limitations to our proposed methodology using BNs, prior knowledge from ontologies, and adaptability and generalizability to other datasets. CBNs are expressed as DAGs, which are acyclic by definition and incapable of expressing any “feedback loops” which are present in complex diseases [52]. Every arc in the model presumes that one event precedes the other; while excluding a potential situation where comorbid events A and B can both be the cause and the effect. For the Bayesian network, the algorithm will choose the best arc direction based on a scoring mechanism and exclude other potentially causal arc directions. However, it is possible to force the calculation of the conditional probabilities for both A->B and B->A given a learned network structure. If we force a direction in a model that is contrary to what is learned from the data, the rest of the model will also be reoriented to compensate for the directional arc changes. Counterfactual studies (the “what-ifs”) of causal models may be included in future research as prior knowledge of potential “causal loops” are presented the AMOs of the disease domain.

The methodology proposed in this paper requires 1) a significantly robust dataset, 2) an AMO with prior causal knowledge, and 3) an algorithm capable of generating a BN with whitelisted or blacklisted arcs which represent prior causal knowledge. Moreover, the prior knowledge presented in the ontology must be useful in identifying or explaining a potential causal mechanism between two variables in the dataset.

II. Extracting and Applying Prior Causal Knowledge From ICD-10-CM

A. Acquiring Prior Causal Knowledge from ICD-10-CM

The ICD-10-CM ontology contains “Other Vital Details”, as well as other object properties for indexed symptom terms. These symptoms are related to other symptoms, which can be used to infer causality between two variables and create an ordered variable pair. ICD-10-CM is a formal hierarchical ontology organized using a straightforward alphanumeric code structure[5]. Since medical coding is not available for any of the forms administered in the NACC UDS, the diagnostic codes are matched manually to the closest ICD-10-CM symptom. All 24 variables in the NACC dataset (excluding age and gender) have been matched, together with notes and the full ICD-10-CM symptom if it is different from the name of the main symptom in Table 3.

Table 3: NACC UDS Form B9 Alzheimer Symptom Variable Names and corresponding ICD-10-CM Codes

Variable Name	Main Symptom(s)	ICD-10-CM Code(s)	Notes
DEMENTED	Dementia	F03	F03 – Unspecified dementia
BEAGIT	Agitation	R45.1	
BEAHALL	Auditory Hallucinations	R44.0	
BEANX	Anxiety	F40, F41	F40 for Phobic Anxiety, F41 for Other

[5] https://www.cdc.gov/nchs/icd/icd10cm.htm

BEAPATHY	Apathy, Withdrawal	R45.3	R45.3 Demoralization and apathy
BEDEL	Delusional Beliefs	F22	F22 for generic delusional disorders
BEDEP	Depression	F32, F33	F32 for single episode, F33 for recurrent
BEDISIN	Disinhibition	R45.87	Impulsiveness R45.87 is an approximation. Disinhibited term occurs under childhood disorders (F94.1), which does not align with the development of AD and is excluded
BEIRRIT	Irritability	R45.4	Irritability and anger R45.4
BEPERCH	Personality Change	F07	Personality and behavioral disorders due to known physiological condition F07
BEREM	Sleep Disorder	G47	
BEVHALL	Visual Hallucinations	R44.1	
COGATTN	Attention, Concentration	R41.840	Attention and concentration deficit R41.840. Attention-deficit hyperactivity disorders (ADHD) F90 is related to childhood and excluded.
COGJUDG	Judgement, Planning, Problem-Solving	R41.844	Frontal lobe and executive function deficit R41.844. Symptom is also under “Frontal lobe and executive function deficit following nontraumatic subarachnoid hemorrhage I69.014” which is excluded.
COGLANG	Language	R47	Speech disturbances R47
COGMEM	Memory	R41.1, R41.2, R41.3	3 types of amnesia: Anterograde R41.1, Retrograde R41.2, and Other R41.3
COGORI	Orientation	R41.0	Disorientation R41.0
COGVIS	Visuospatial Function	R41.842	Visuospatial deficit R41.842
MOFALLS	Falling	R29.6	Repeated falls R29.6
MOGAIT	Gait Disorder	R26	Abnormalities of gait and mobility R26
MOMOALS	Lateral Sclerosis	G12.21, G12.23	Primary lateral sclerosis G12.23, Amyotrophic lateral sclerosis (ALS) G12.21. While variable name suggests ALS, no clarification is given in NACC UDS data dictionary
MOMOPARK	Parkinson	G20	
MOSLOW	Slowness	R46.4	Slowness and poor responsiveness R46.4
MOTREM	Tremor	R25.1	Tremor, unspecified R25.1

Many of the NACC UDS variables have direct matches to ICD-10-CM definitions. For example, Parkinson’s Disease (MOMOPARK), Sleep Disorder (BEREM), and Delusional Beliefs (BEDEL) all have dedicated ICD-10-CM classifications. Other symptoms have precise translations when codified:

- Impairment of memory (COGMEM) is matched to different types of amnesia (R41.1, R41.2, R41.3)
- Impairment of language (COGLANG) is matched to speech disturbances (R47)
- Symptoms without direct matches are approximated.

Question 4C on Form B9 assesses executive function: “Executive function — judgment, planning, problem-solving. Does s/he have trouble handling money (e.g., tips), paying bills, preparing meals, shopping, using appliances, handling medications, driving?” The resulting variable COGJUDG focuses on multiple main symptoms (judgment, planning, problem-solving) which are not codified in ICD-10-CM or other AMOs. Executive function, however, is listed but does not specify context for the symptoms experienced.

Question 9C in form B9 for Disinhibition (BEDISIN) asks: “Does the subject use inappropriate coarse language or exhibit inappropriate speech or behaviors in public or in the home? Does s/he talk personally to strangers or have disregard for personal hygiene?” The main symptom of disinhibition is classified under “Disorders of social functioning with onset specific to childhood and adolescence” (F94) as Disinhibited attachment disorder of childhood (F94.1). However, since AD does not develop until adulthood[6], this classification is excluded from the experiment. Finally, disinhibition

[6] https://www.nia.nih.gov/health/what-are-signs-alzheimers-disease

(BEDISIN) and the symptoms described under the question is approximated as Impulsiveness (R45.87). Using diagnostic clinical information in the https://www.icd10data.com browser, we determine that the following 10 causal relationships potentially exist among the 24 variables:

Table 4: Ordered Variable Pairs for NACC UDS CBN from ICD-10-CM

Edge	**Notes on Causality from ICD-10-CM**
COGATTN → BEDISIN	Impulsivity is a main feature of ADHD, along with inattention and hyperactivity.
COGORI → BEVHALL	Disorientation is characterized by distortions of visual and auditory perceptions, including illusions and hallucinations
COGORI → BEAHALL	Disorientation is characterized by distortions of visual and auditory perceptions, including illusions and hallucinations
DEMENTED → COGLANG	Dementia causes problems in two or more brain functions, such as memory and language
DEMENTED → COGMEM	Dementia causes problems in two or more brain functions, such as memory and language
DEMENTED → COGJUDG	Dementia affects executive judgment
MOGAIT → MOFALLS	Irregular gait causes stumbling and falls
MOMOALS → MOFALLS	Tripping and balance issues are caused by ALS as leg muscles weaken
MOMOPARK → MOTREM	Parkinson's is a progressive motor disability which causes tremors, shaking, and muscular rigidity
MOMOPARK → MOGAIT	Parkinson's causes postural instability, and gait abnormalities. This is due to a loss of neurons and a decrease of dopamine in the basal ganglia

B. Applying Prior AMO Causal Knowledge using Min-Max Hill Climbing Algorithm (MMHC)

Applying prior expertise of causal knowledge is possible by identifying and applying variable pairs, as we previously established in [67]. We identified the MMHC algorithm as being one of the most accurate in terms of log-loss score and will use it to learn both a baseline BN and an improved CBN incorporating the variable pairs in Table 4. MMHC is comprised of two algorithms in stages:

Stage 1: The Max-Min Parents & Children (MMPC) [47] algorithm is used to learn the parents and children nodes of a variable X (*PCx*), as a subset of the variables (*V)* in the data (*D)*.

Stage 2: A greedy hill-climbing algorithm is applied to the *PCx* discovered using MMPC. A random symptom *Y* is selected, and neighboring symptoms are searched for associations using the Dirichlet likelihood-equivalence uniform score (BDeu) [47] within *PCx*. The BDeu score aims at maximizing the posterior probability of the directed acyclic graph (DAG) based on the dataset, while assuming a uniform prior distribution over possible DAGs.

During the second stage of the MMHC algorithm, *the conditions in which the* `add-edge` *operator is modified to consider prior ordered pairs*. Previously, an edge is added if *Y* ∈ *PCx*. For the algorithm to consider the ordered variable pairs, the conditions for using the `add-edge` operator within MMHC must now satisfy two conditions:

1. *Y* ∈ *PCx* (this is an original condition for the MMHC algorithm, and will remain); and
2. *Y* to *X* does not violate the direction of any ordered variable pairs previously discovered in the AMO.

Condition 1) remains from the original MMHC algorithm, and states that edge *Y* to *X* may only be added if it exists in *PCx*. Condition 2) is an addition to the MMHC algorithm. This modification checks the edge addition against a list of previously established ordered variable pairs. The pseudocode for the modified algorithm is as follows:

```
Procedure MMHC with Ordered Variable Pairs(D)
Input: dataset D with variable set V
Output: DAG based on the variables in D
For every variable X ∈ V: PCx = MMPC(X,D)
     Start with an empty graph and perform greedy hill-climb (add/delete/reverse edge).
```

```
    Only try add-edge operator for Y to X if Y ∈ PCx AND if Y to X conforms to a list of
    apriori ordered variable pairs
Return the highest scoring DAG
End procedure
```

For *Y* to *X* to consider prior knowledge in the form of ordered pairs, the edge direction of *Y* to *X* cannot violate a previously established ordered pair. For example, if a prior sequence suggests that a directed edge exists from *X* to *Y*, then the edge *Y* to *X* will not be added by the greedy hill-climbing algorithm, regardless of whether $Y \in PCx$ or if there is a high BDeu score for *Y* to *X*. The pseudocode for the modified greedy hill-climbing algorithm with previously ordered variable pairs is as follows:

```
Procedure Greedy HC with Ordered Variable Pairs (PCx)
Input: Parent-Children sets from MMPC
Output: An DAG X with highest overall BDeu Score
Start with an empty graph
For every PCx graph:
   add-edge Y to X if Y ∈ PCx AND if Y to X  conforms to a list of apriori ordered
   variable pairs
       delete-edge Y from X if higher score from    PCx exist
       reverse-edge Y from X if a reversal results   in a  higher score
Return highest scoring DAG X
End procedure
```

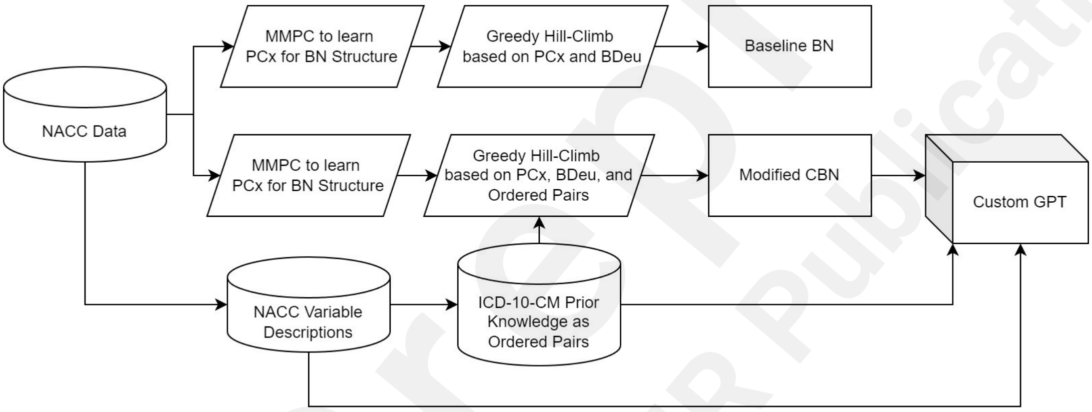


Fig. 4: Process for Creating a Baseline and Modified CBNs

Since these modifications alter the structure of the learned Bayesian network, the conditional probability parameters of the edges in the network will also change. The changes depend on which edges have been generated, and the conditional dependencies that are formed within the network. A baseline CBN model without prior ordering can be generated by applying the unmodified MMHC algorithm. An unmodified MMHC algorithm in the *bnlearn* package in *R* can be used to learn the structure of the network in conjunction with the *bn.fit* function to learn the parameters of the structure.

C. Comparative Analysis of Baseline vs. Modified CBNs

Modifying the algorithm with prior ordered variable pairs removes the possibility for the algorithm to learn any arcs which contradicts prior ontological knowledge. Furthermore, since the arcs are oriented based on prior pairs, new BDeu scores are calculated resulting in new relationships being formed. Previously, these ten ordered variable pairs are translated into a blacklist for the modified CBN:

```
blist = data.frame(
```

```
from=c("BEDISIN","BEVHALL","BEAHALL","COGLANG","COGMEM","COGJUDG"
,"MOFALLS","MOFALLS","MOTREM","MOGAIT"))
to=c("COGATTN","COGORI","COGORI","DEMENTED","DEMENTED","DEMENTED"
,"MOGAIT","MOMOALS","MOMOPARK","MOMOPARK"),
```

The Modified CBN model is learned with the blacklist (`blist`) as prior knowledge:

```
netsmart2 = mmhc(x=df, whitelist = NULL, blacklist = blist)
```

Finally, we will blacklist reversed AMO variable pairs and examine the agreement of arcs, and area under the ROC curve (AUC) for MMHC and regular HC, between Baseline (netb) and Modified (netm) models.

```
 > compare(netb,netm)
$tp       $fp       $fn
[1] 42    [1] 7     [1]
                    8
```

With the following false positive (`fp`) and false negative (`fn`) arcs:

```
$fp                             $fn
     from       to                   from       to
[1,] "MOMOPARK" "MOGAIT"        [1,] "MOGAIT"
[2,] "MOTREM"   "MOSLOW"        "MOMOPARK"
[3,] "GENDER"   "MOMOPARK"      [2,] "MOMOPARK" "GENDER"
[4,] "MOGAIT"   "MOFALLS"       [3,] "MOFALLS"  "MOTREM"
[5,] "MOTREM"   "MOFALLS"       [4,] "MOSLOW"   "MOTREM"
[6,] "GENDER"   "MOTREM"        [5,] "MOTREM"   "GENDER"
[7,] "MOMOALS"  "MOMOALS"       [6,] "BEDEP"    "GENDER"
                                [7,] "COGVIS"
                                "MOFALLS"
                                [8,] "MOFALLS"  "MOGAIT"
```

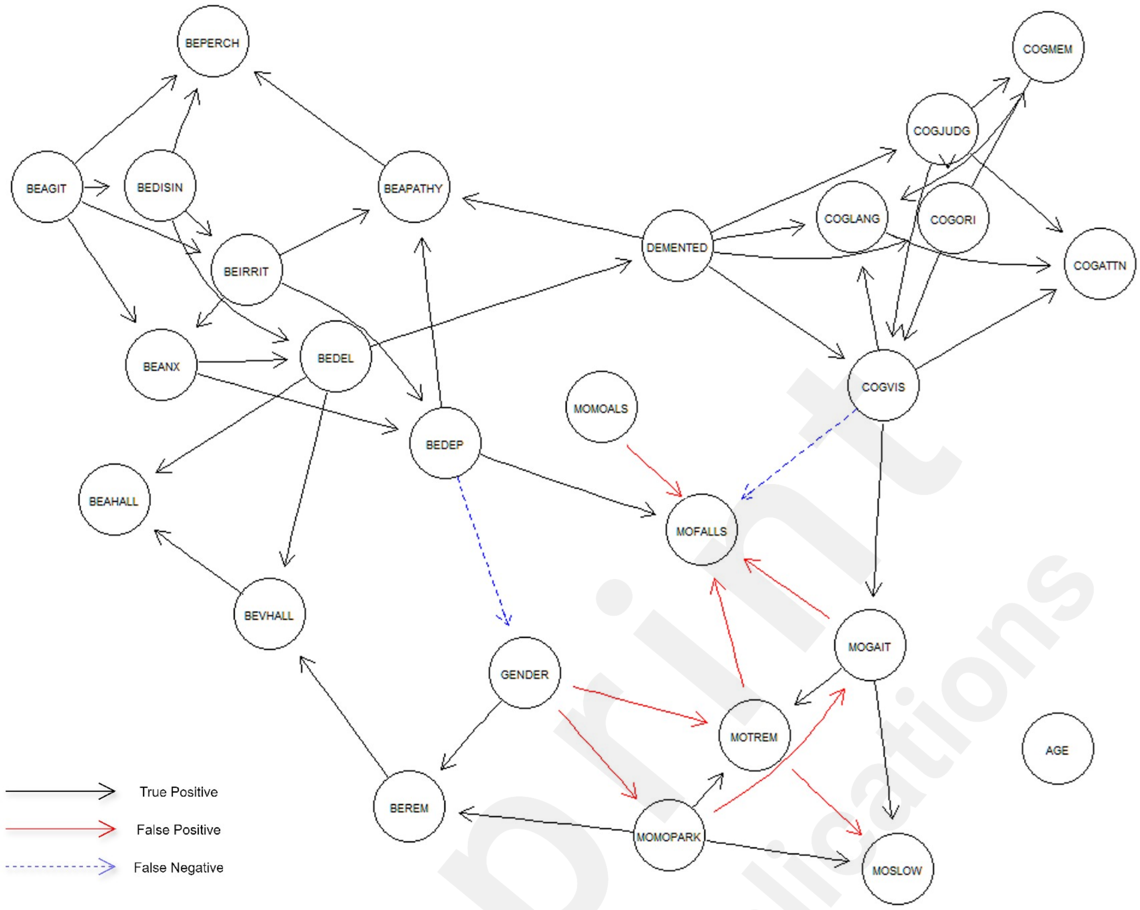


Fig. 5: Arc Agreement between Baseline vs. Modified CBN

Blacklisting the AMO ordered variable pairs, the Modified network has a 41/48 (85%) agreement. The AUC for Baseline `netb` vs. the `netm` Modified CBN:

```
MMHC Baseline AUC: 0.9514, 95% CI: (0.9386, 0.9523)
MMHC Modified AUC: 0.9617, 95% CI: (0.9511, 0.9634)
HC Baseline AUC: 0.8586, 95% CI: (0.8388, 0.8711)
HC Modified AUC: 0.8831, 95% CI: (0.8652, 0.9043)
```

Adding ontological constraints as blacklist created a Modified network which is marginally better than the Baseline in terms of predictive accuracy. Hill Climbing (HC) improved slightly more due to the ceiling effect of how well MMHC can perform. On average, the MMHC Modified CBN had an increase of 1.08%, and regular HC had an increase of 2.85%. The added ontological constraints increased congruence between model and ontological causal claims, indicating that the modified network agrees with both the ICD-10-CM ontology and the potentially causal relationships discovered in the NACC data.

MMHC Baseline AUC: 0.9514 95% CI: (0.9386, 0.9523)	MMHC Modified AUC: 0.9592 95% CI: (0.9521, 0.9664)

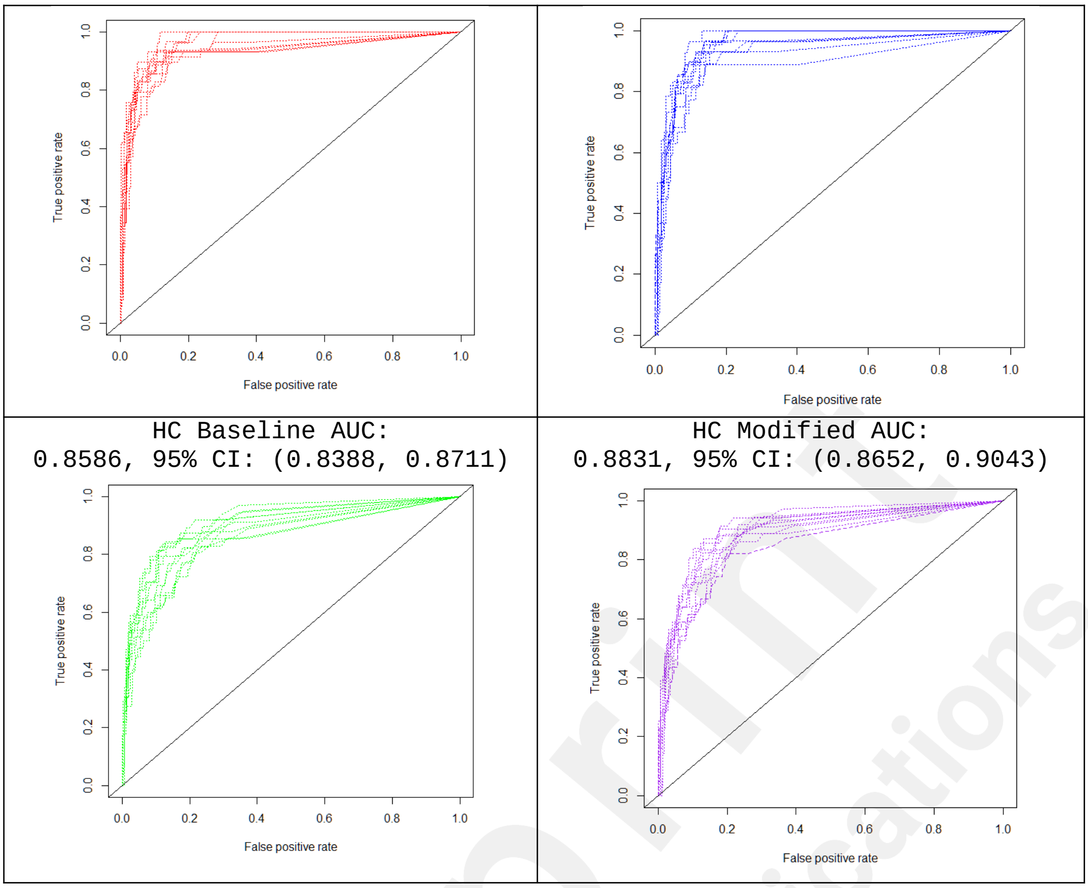


Fig. 6: Cross-validated AUC for Baseline vs. Modified NACC CBNs

III. A Custom GPT Advisor for Alzheimer's Disease

A. Constructing a Custom GPT for Alzheimer's Disease

In November 2023, OpenAI's introduced the ability to define custom GPTs[7] allowing users to tailor ChatGPT for specific tasks, including acting as a medical recommender system [76]. These customizations can be complex, to include specific causal information and conditional probabilities in a CBN. Custom GPTs extend the existing knowledge of ChatGPT-4O with additional instructions, data, and conditional statements. The additional parameters can be added to Custom GPTs via GPT Builder, which is able to process text and uploaded files and images.

For our GPT Recommender System for Alzheimer's Disease, we will provide the GPT Builder[8] with the following information: 1) Conditional probabilities among comorbid symptom variables in the Modified CBN, 2) The descriptions of NACC Variables and their abbreviations, and 3) ICD-10-CM ordered variable pairs indicating potentially causal relationships among NACC Variables. Additionally, we will also instruct the GPT to respond as a clinician and to not provide responses that are not pertinent to Alzheimer's Disease.

[7] https://openai.com/index/introducing-gpts/
[8] https://help.openai.com/en/articles/8770868-gpt-builder

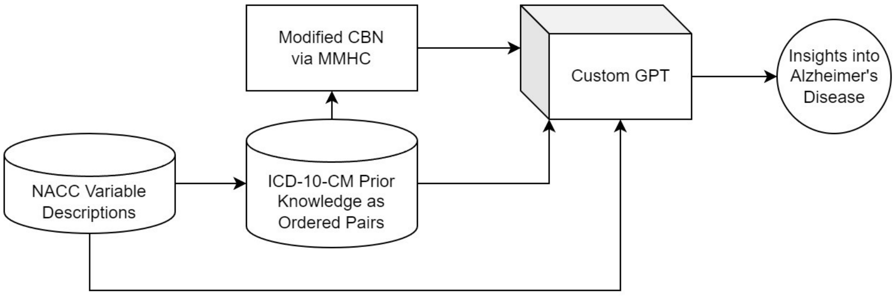


Fig. 7: Sources of Specialized Information for GPT Recommender System

Entering the new information in our custom GPT via GPT Builder is simple, and the system can either receive the information in text or uploaded file formats. We will upload the Modified CBN relationships and conditional probabilities, NACC variables and variable descriptions, and ICD-10-CM codes and ordered variable pairs as a delimited CSV file.

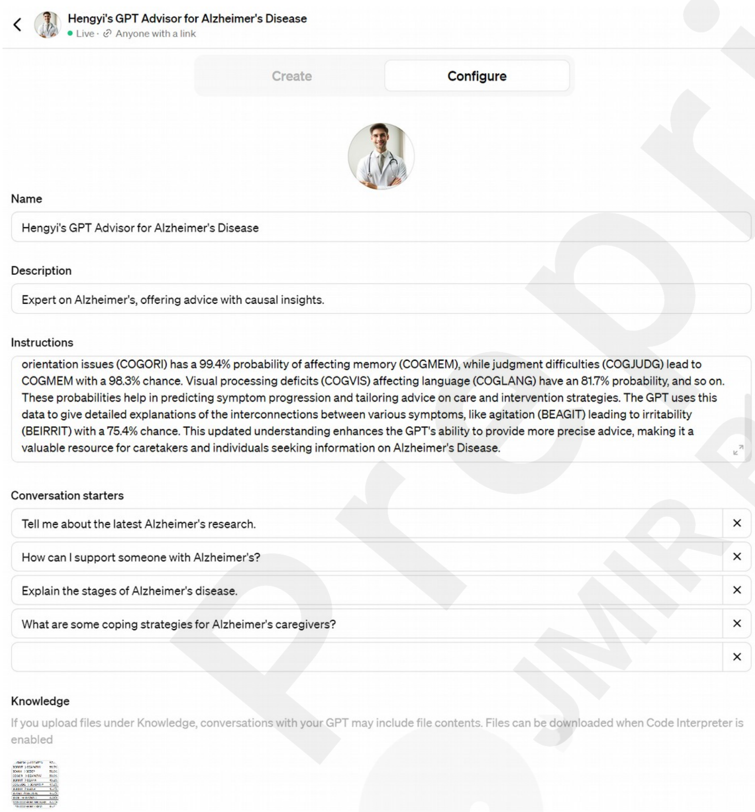


Fig. 8: GPT Builder for Custom GPT Recommender System

B. Causal Insights into Alzheimer's Disease using a Custom GPT

To assess the causal insights provided by the custom GPT, we will 1) compare ChatGPT-4Oo to our custom GPT Advisor, and 2) qualitatively explore potentially causal relationships in existing epidemiological literature. While we are able to perform comparative analysis on our Baseline and Modified CBNs, we will not be able to assess the accuracy of causal statements made by ChatGPT-4Oo because it is based on text patterns, and not an analysis of conditional probabilities found in existing clinical data [77]. To compare ChatGPT-4Oo with our custom GPT, we will submit a couple of general queries based on the NIH's National Institute on Aging's most frequently asked questions[9],

[9] https://www.nia.nih.gov/health/alzheimers-and-dementia/8-frequently-asked-questions-about-

and a several specific queries designed to gather insights into symptom comorbidity.

Q1: What is the difference between Alzheimer's disease and dementia?	
ChatGPT-4Oo: **Dementia** is an umbrella term used to describe a range of symptoms affecting memory, thinking, and social abilities severely enough to interfere with daily life. Various conditions can cause dementia, including Alzheimer's disease, vascular dementia, Lewy body dementia, and frontotemporal dementia. **Alzheimer's Disease** is the most common cause of dementia, accounting for 60-80% of cases. It is a specific disease that causes the brain to shrink (atrophy) and brain cells to die.	Custom GPT: **Dementia** is a general term that describes a decline in cognitive function severe enough to interfere with daily life. It is not a specific disease but an umbrella term for symptoms affecting memory, communication, and thinking. Various conditions can cause dementia, including Alzheimer's disease, vascular dementia, Lewy body dementia, and frontotemporal dementia. **Alzheimer's disease** is the most common cause of dementia, accounting for 60-80% of cases. It is a specific disease characterized by the progressive degeneration of brain cells, leading to cognitive decline, memory loss, and changes in behavior and personality.
Q2: What are the comorbid symptoms in AD, in one paragraph?	
ChatGPT-4Oo: Comorbid symptoms in Alzheimer's Disease (AD) significantly impact the patient's overall health and quality of life. Common comorbidities include cardiovascular diseases such as hypertension and diabetes, which can exacerbate cognitive decline and complicate treatment. Psychiatric conditions like depression and anxiety are also prevalent, affecting nearly half of all Alzheimer's patients, and can lead to worsened cognitive and functional impairment. Additionally, sleep disturbances and sensory impairments, such as vision and hearing loss, are frequent and contribute to increased confusion and reduced quality of life. Managing these comorbidities requires a comprehensive and integrated approach to care, addressing both cognitive and physical health needs	Custom GPT: In Alzheimer's Disease (AD), comorbid symptoms often accompany the primary cognitive decline, significantly impacting overall health and behavior. Common cognitive symptoms include memory impairment, orientation issues, and judgment difficulties, which frequently interrelate. Behavioral symptoms such as agitation, depression, and apathy are prevalent, often exacerbating one another. Psychological symptoms like anxiety and paranoia are also common, contributing to increased behavioral disturbances. Physical symptoms, including sleep disturbances and appetite changes, further complicate the clinical picture. Visual processing deficits and language difficulties are additional challenges, interlinked with cognitive symptoms and affecting daily functioning. Understanding these comorbidities helps in predicting disease progression and optimizing care strategies.

Q1 and Q2 are generic questions regarding AD, with GPT-4Oo offering a more generic response and the Custom GPT naming specific symptoms related to AD. In Q2, the Custom GPT provides a more detailed answer to include specific psychological and physiological symptoms that were a part of the dataset, and their implications as comorbid symptoms.

Q3: Tell me about the cognitive impairments of AD and how it affects other AD symptoms	
ChatGPT-4Oo: Alzheimer's Disease (AD) primarily manifests through progressive cognitive impairments, which significantly impact other symptoms and overall functioning. The main cognitive impairments in AD include memory loss, particularly affecting recent events and new information, which is often one of the earliest and most recognizable symptoms. As the disease progresses, patients experience difficulties with language, such as finding the right words or following conversations, and visuospatial skills, leading to challenges in navigation and recognizing familiar faces and objects. These cognitive impairments exacerbate other AD symptoms, such as behavioral and psychological symptoms. For instance, memory loss can lead to increased anxiety, confusion, and frustration, which may result in agitation or aggression. The decline in executive function and judgment can lead to poor decision-making and unsafe behaviors, increasing the risk of accidents. Additionally, impairments in attention and concentration can make it difficult for patients to engage in daily activities, leading to a loss of independence and increased caregiver burden.	Custom GPT: Cognitive impairments in Alzheimer's Disease (AD) primarily involve memory loss, difficulties with orientation, poor judgment, and diminished problem-solving abilities. These cognitive deficits profoundly affect other symptoms of AD: 1. **Memory Impairment (COGMEM)**: As memory declines, patients struggle to remember recent events, recognize familiar faces, or recall names. This memory loss leads to confusion and disorientation, contributing to a decline in orientation (COGORI) with a 99.4% probability. 2. **Orientation Issues (COGORI)**: Difficulty in understanding time, place, and identity further disrupts daily activities and increases reliance on caregivers. It exacerbates memory issues and contributes to poor judgment and decision-making. 3. **Judgment and Problem-Solving Difficulties (COGJUDG)**: Impaired judgment leads to unsafe decisions and increased risk of accidents. There's a 98.3% chance that poor judgment will further impair memory. These difficulties affect the ability to manage finances, medication, and daily routines, increasing dependence on others. These cognitive impairments also affect behavioral and psychological symptoms. For example, poor judgment and memory loss can lead to frustration, resulting in increased agitation and irritability. The confusion from orientation issues can cause anxiety and depression, further complicating the behavioral health of the

alzheimers-disease

	patient. Cognitive decline also impacts communication skills, leading to social withdrawal and isolation, which can worsen depression and apathy.

Q3 specifically asks a question of cognitive impairment, which is a classification of symptoms, and how it relate to – and affects – other AD symptoms. The answers for specific queries vary greatly between the two GPTs, with the Custom GPT speaking directly to potentially causal relationships found in the CBN. The response also cites specific conditional probabilities found in the dataset, then extends this conditional probability with explanations which are external to the CBN.

Q4: Tell me how Memory Loss relates to other AD symptoms, in a paragraph	
<u>ChatGPT-4Oo:</u> Memory loss is a hallmark symptom of Alzheimer's Disease (AD) and is intricately connected to other cognitive and behavioral symptoms. As the disease progresses, memory loss typically starts with short-term memory lapses, making it difficult for individuals to remember recent events or conversations. This impairment extends to other cognitive functions such as language, leading to difficulties in finding the right words or understanding conversations. Memory loss also disrupts spatial awareness, causing disorientation and difficulties in navigating familiar environments. Additionally, the decline in executive function, which involves planning, decision-making, and problem-solving, is exacerbated by memory deficits. Behavioral symptoms like agitation, anxiety, and depression often arise as individuals struggle with their diminishing cognitive abilities. Therefore, memory loss in AD is not an isolated issue but a central symptom that exacerbates and intertwines with other aspects of the disease's progression.	<u>Custom GPT:</u> Memory loss in Alzheimer's Disease (AD) is intricately linked to other symptoms, significantly affecting overall cognitive, behavioral, and functional abilities. There's a 99.4% probability that memory loss will worsen orientation issues, leading to confusion about time, place, and personal identity. It also has a 98.3% chance of impairing judgment and decision-making, causing poor problem-solving and unsafe decisions. Visual processing deficits, which can influence memory, have an 81.7% probability of affecting language abilities, making communication challenging. Memory loss contributes to behavioral symptoms such as agitation and irritability, with a 75.4% probability that agitation will increase irritability. These cognitive impairments exacerbate psychological symptoms like depression and anxiety, further complicating daily life. Understanding these probabilities and interconnections is crucial for comprehensive management of AD.

Q4 asks about a singular symptom (memory loss), with Custom GPT summarizing causal relationships into a single paragraph. This paragraph describes a causal sequence of events among all known relationships and conditional probabilities associated with memory loss, and how it affects other symptoms as a result. In the unmodified ChatGPT-4Oo, several comorbid symptoms are listed which are absent in the Custom GPT. These symptoms imply causation, e.g. "memory loss leads to depression", which is not observed in the CBN. While this description qualitatively makes sense, there is simply no conditionally independent path between BEDEP and COGMEM in the CBN.

Q5: Tell me about the relationship between attention deficit and disinhibition in a paragraph	
<u>ChatGPT-4Oo:</u> Attention deficit and disinhibition are closely linked cognitive phenomena often observed in conditions such as Attention Deficit Hyperactivity Disorder (ADHD). Attention deficit refers to the inability to sustain focus on tasks, leading to distractibility and difficulties in completing activities. Disinhibition, on the other hand, involves a lack of restraint manifested in impulsive behavior, poor judgment, and a propensity to act without considering consequences. Both can result from dysfunctions in the brain's executive functions, particularly within the prefrontal cortex. The interplay between these two aspects can exacerbate each other; for instance, the inability to maintain attention can lead to impulsive reactions, while impulsivity can further disrupt focus. This relationship complicates daily functioning and often requires integrated therapeutic approaches to manage effectively.	<u>Custom GPT:</u> In Alzheimer's disease, attention deficits (COGATTN) and disinhibition (BEDISIN) are interconnected symptoms. Attention deficits impair the ability to maintain focus and process information, which disrupts executive functions essential for impulse control and social appropriateness. This impairment increases the likelihood of disinhibited behavior, such as impulsivity and inappropriate actions, due to weakened cognitive control. The model indicates a 62.1% probability that attention deficits will lead to disinhibition, highlighting a significant risk of behavioral changes as the disease progresses. Understanding this relationship helps caregivers and healthcare providers implement strategies to manage symptoms and improve patient care.

Q5 specifically asks for the relationship between two symptoms in the dataset. Without mentioning AD, GPT-4Oo responds generally with an answer that is not related to AD. The Custom GPT responds within the context of AD, as well as a directional causal relationship (COGATTN → BEDISIN) from the CBN with associated conditional probability. Lastly, to assess the relevance and accuracy of our CBN and custom GPT responses, we examine potentially causal relationships within our CBN using existing epidemiological literature:

- COGATTN → BEDISIN: An inability to concentrate is a potential cause of cognitive and behavioral disinhibition, as patients loses the ability to regulate emotion. Attention deficits impair the ability to focus and process information, which disrupts executive functions such as impulse control and social appropriateness. This disruption increases the likelihood of disinhibited behaviors [78], [79].

- MOGAIT → MOFALLS: The decline in frontal cognitive functions contributes to alterations of gait. An altered gait significantly increases the risk of falls in AD patients, as gate and balance are precursors to falls [80].

- MOMOALS → MOFALLS: Amyotrophic Lateral Sclerosis (ALS), also known as Lou Gehrig's disease, encompasses a range of issues, including difficulty walking, reduced muscle strength, and impaired coordination, all of which contribute to instability and a higher likelihood of falling [81].

- COGORI → BEVHALL and COGORI → BEAHALL: Cognitive orientation (COGORI) is a very strong indicator of auditory hallucinations (BEAHALL, 94%) and visual hallucinations (BEVHALL, 93%) in Alzheimer patients. These two relationships appear in the Modified CBN but are not present in the Baseline model. Current epidemiological studies show that:
  - Hallucinations and delusions are symptoms associated with mid-late stage Alzheimer's disease [82], [83], [84], while disorientation and special navigation is one of the earliest symptoms of the disease. [85], [86], [87];
  - Disorientation resulting from cognitive orientation issues can contribute to these hallucinations by creating confusion and a disjointed perception of reality. Disorientation for patients of AD include impaired perception of surroundings, which relates to the "sequence of early hippocampal and later posterior cortical damage that is typical of AD" [88], [89]; and
  - The progression of cognitive decline and disorientation leads to visual and auditory disorders. [90], [91], [92], [93].

- MOMOPARK → MOGAIT: Distinctive symptoms of Parkinson's Disease include irregular gait, slowness, and tremor [94].

- BEVHALL → BEDEL and BEAHALL → BEDEL: Hallucinations (visual and auditory) are associated with delusions and the comorbid symptoms for AD patients exacerbate each other over the course of the disease [83], [84], [90].

- BEDISIN → BEAGIT: Disinhibition and agitation are both categorized as the inability to control and regulate emotion, and disinhibited behaviors can often lead to increased agitation. A patient's impulsive actions may create situations that are stressful or confusing, provoking a heightened state of irritability and agitation [95].

- GENDER → MOMOPARK: Parkinson's Disease is 1.5 times more common in males than females [96].

- COGMEM → COGORI: Memory and orientation issues are preliminary symptoms to AD [97].

- DEMENTED (dementia) in the context of AD is the potential of several comorbid symptoms

[93], [97], [98], including:

- DEMENTED → COGORI (Orientation Issues): 97.8%, Dementia frequently leads to difficulties with orientation, such as disorientation to time, place, or person.
- DEMENTED → COGJUDG (Judgment Issues): 98.3%, Cognitive judgment issues are highly likely to occur in individuals with dementia, affecting decision-making and problem-solving abilities.
- DEMENTED → COGVIS (Visual Processing Issues): 81.7%, Dementia is associated with a high probability of visual processing deficits, impacting the ability to interpret visual information correctly.
- DEMENTED → COGLANG (Language Issues): 85.6%, Language problems, including difficulties with finding words and constructing sentences, are common in dementia.

C. Conclusions

ChatGPT and LLMs are trained on text data, which primarily reflects correlations rather than causal relationships. While the model learns patterns and associations within the training text, it does not inherently understand any underlying causal mechanisms. It recognizes that certain events often occur together but does not know why one event causes another. This is because LLMs like ChatGPT generate responses based on the likelihood of word sequences as determined by their training data. While this probabilistic approach generates coherent and contextually relevant text, it does not inherently understand cause-and-effect, or the fact that a cause precedes the effect. Causation requires an understanding of “how” and “why” one event affects another, which is not considered in pattern recognition. The inability of generative AI to reliably understand cause-and-effect has significant implications across various domains. In healthcare, this limitation can lead to incorrect diagnoses or ineffective treatment recommendations, as the AI might fail to consider the causal relationships between symptoms and underlying conditions.

Establishing causation requires controlled experiments and observations. ChatGPT, as of May 2024, does not have access to experimental data or the ability to learn models from data. It relies solely on pre-existing textual correlations. Without specific knowledge of cause and effect, LLMs are unable to identify the underlying causes of symptoms or quantify their relationships to other comorbid symptoms. A LLM might correlate certain symptoms with a particular disease without recognizing that these symptoms are secondary to another primary condition. Symptoms such as Memory Loss, show up earlier in course of the Alzheimer’s Disease and is a known cause of Loss of Orientation and Loss of Executive Judgment. Since LLMs are unable to inherently grasp causal relationships and may also be limited by the amount of available data, they does not provide an effective answer in predicting disease progression and patient outcomes.

In clinical settings, decision-support and recommender systems that suggest causal relationships among comorbid symptoms or recommend interventions must have direct causal understanding to suggest appropriate treatments. Without the ability to reliably discern cause and effect undermines the extended benefits of generative AI and LLMs in healthcare. Currently imbuing existing ChatGPT models with external data analysis creates a more complete GPT model that can be used to interpret and leverage causal relationships to improve patient care and medical research. Creating a custom GPT also makes knowledge in causal network models accessible to patients, clinicians, and researchers. In the absence of direct causal information, the custom GPT can still produce useful information and citations about AD in general.

To obtain our CBN, we’ve used a previous methodology which obtains ordered variable pairs from authoritative ontology ICD-10-CM as prior expertise. In our past research, a CBN that is learned using MMHC can be improved significantly by considering prior sources of knowledge, if the algorithm is modified appropriately. This source of prior knowledge can be validated in existing literature as a sequence of events in AD progression, specific causal mechanisms among comorbid symptoms, and conditional probabilities from a Bayesian Network. The resulting modified network provides insight into the causal relationships expressed in the AD data and takes advantage of the expertise and knowledge contained in the AMOs.

Since inferring causality from a CBN does not exist in a vacuum, the relationships within the CBN, regardless of the strength of the conditional probabilities, must be explored further. A custom GPT extends the causal knowledge within a CBN with general knowledge of symptoms and diseases, providing a tool capable of suggesting causal inference based on the analysis of real patient data. With a LLM, uncertainty is present but reasoned with, as it is present but reasoned with in the prior, posterior, and likelihood in Bayes Theorem. Moving forward, we would like to explore using ChatGPT to produce ordered variable pairs and to automate the validation of potentially causal information.

References

[1] E. C. Garrido-Merchán *et al.*, "Real Customization or Just Marketing: Are Customized Versions of Chat GPT Useful?" arXiv, Nov. 27, 2023. doi: 10.48550/arXiv.2312.03728.

[2] L. Villa, D. Carneros-Prado, A. Sánchez-Miguel, C. C. Dobrescu, and R. Hervás, "Conversational Agent Development Through Large Language Models: Approach with GPT," in *Proceedings of the 15th International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2023)*, J. Bravo and G. Urzáiz, Eds., Cham: Springer Nature Switzerland, 2023, pp. 286–297. doi: 10.1007/978-3-031-48306-6_29.

[3] L. A. Cox Jr., "Pushing Back on AI: A Dialogue with ChatGPT on Causal Inference in Epidemiology," in *AI-ML for Decision and Risk Analysis: Challenges and Opportunities for Normative Decision Theory*, L. A. Cox Jr., Ed., in International Series in Operations Research & Management Science. , Cham: Springer International Publishing, 2023, pp. 407–423. doi: 10.1007/978-3-031-32013-2_13.

[4] R. Tu, C. Ma, and C. Zhang, "Causal-Discovery Performance of ChatGPT in the context of Neuropathic Pain Diagnosis," *arXiv e-prints*. Jan. 01, 2023. doi: 10.48550/arXiv.2301.13819.

[5] J. Gao, X. Ding, B. Qin, and T. Liu, "Is ChatGPT a Good Causal Reasoner? A Comprehensive Evaluation." arXiv, Oct. 12, 2023. doi: 10.48550/arXiv.2305.07375.

[6] R. Rokey, L. A. Rolak, Y. Harati, N. Kutka, and M. S. Verani, "Coronary artery disease in patients with cerebrovascular disease: A prospective study," *Ann. Neurol.*, vol. 16, no. 1, pp. 50–53, 1984, doi: https://doi.org/10.1002/ana.410160110.

[7] C. A. Roldan, E. A. Gelgand, C. R. Qualls, and W. L. Sibbitt, "Valvular Heart Disease as a Cause of Cerebrovascular Disease in Patients With Systemic Lupus Erythematosus," *Am. J. Cardiol.*, vol. 95, no. 12, pp. 1441–1447, Jun. 2005, doi: 10.1016/j.amjcard.2005.02.010.

[8] F. Wolfe, B. Freundlich, and W. L. Straus, "Increase in cardiovascular and cerebrovascular disease prevalence in rheumatoid arthritis.," *J. Rheumatol.*, vol. 30, no. 1, pp. 36–40, Jan. 2003.

[9] M. Zong and B. Krishnamachari, "Solving math word problems concerning systems of equations with GPT models," *Mach. Learn. Appl.*, vol. 14, p. 100506, Dec. 2023, doi: 10.1016/j.mlwa.2023.100506.

[10] S. Wolfram, *What Is ChatGPT Doing ... and Why Does It Work?* Champaign, Illinois: Wolfram Media Inc, 2023.

[11] M. Lee, "A Mathematical Investigation of Hallucination and Creativity in GPT Models," *Mathematics*, vol. 11, no. 10, Art. no. 10, Jan. 2023, doi: 10.3390/math11102320.

[12] L. M. Titus, "Does ChatGPT have semantic understanding? A problem with the statistics-of-occurrence strategy," *Cogn. Syst. Res.*, vol. 83, p. 101174, Jan. 2024, doi: 10.1016/j.cogsys.2023.101174.

[13] X. Huang, D. Estau, X. Liu, Y. Yu, J. Qin, and Z. Li, "Evaluating the performance of ChatGPT in clinical pharmacy: A comparative study of ChatGPT and clinical pharmacists," *Br. J. Clin. Pharmacol.*, vol. 90, no. 1, pp. 232–238, 2024, doi: 10.1111/bcp.15896.

[14] J. Pearl, M. Glymour, and N. P. Jewell, *Causal Inference in Statistics: A Primer*, 1 edition. Chichester, UK ; Hoboken, NJ: Wiley, 2016.

[15] B. Bulik-Sullivan *et al.*, "An atlas of genetic correlations across human diseases and traits," *Nat. Genet.*, vol. 47, no. 11, Art. no. 11, Nov. 2015, doi: 10.1038/ng.3406.

[16] M. Georges, "The long and winding road from correlation to causation," *Nat. Genet.*, vol. 43, no. 3, Art. no. 3, Mar. 2011, doi: 10.1038/ng0311-180.

[17] W. C. Salmon, *Causality and Explanation*. Oxford University Press, 1998.

[18] A. Michotte, *The Perception of Causality*. Routledge, 2017.

[19] M. M. Marini and B. Singer, "Causality in the Social Sciences," *Sociol. Methodol.*, vol. 18, pp. 347–409, 1988, doi: 10.2307/271053.

[20] M. Ramachandran, "A counterfactual analysis of causation," *Mind*, vol. 106, no. 422, pp. 263–277, Apr. 1997, doi: 10.1093/mind/106.422.263.

[21] J. R. Anderson and C.-F. Sheu, "Causal inferences as perceptual judgments," *Mem. Cognit.*, vol. 23, no. 4, pp. 510–524, Jul. 1995, doi: 10.3758/BF03197251.

[22] "Streetwise: Race, Class, and Change in an Urban Community | Sociology." Accessed: Aug. 11, 2020. [Online]. Available: https://sociology.yale.edu/publications/streetwise-race-class-and-change-urban-community

[23] H. L. Smith, “Effects of Causes and Causes of Effects: Some Remarks From the Sociological Side,” *Sociol. Methods Res.*, vol. 43, no. 3, pp. 406–415, Aug. 2014, doi: 10.1177/0049124114521149.
[24] S. Wright, “Correlation and Causation,” *J Agric Res*, vol. 20, pp. 557–585, 1921.
[25] J. Pearl, *Graphs, Causality, And Structural Equation Models*. 1998.
[26] N. Fenton, M. Neil, and M. Neil, *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press, 2012. doi: 10.1201/9780367803018.
[27] N. Friedman, “Inferring cellular networks using probabilistic graphical models,” *Science*, vol. 303, no. 5659, pp. 799–805, Feb. 2004, doi: 10.1126/science.1094068.
[28] J. Pearl, *Causality*. Cambridge University Press, 2009.
[29] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.
[30] J. Pearl, “Bayesian networks,” *UCLA Dep. Stat. UCLA*, Oct. 2011, Accessed: Mar. 19, 2019. [Online]. Available: https://escholarship.org/uc/item/53n4f34m#author
[31] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian Network Classifiers,” *Mach. Learn.*, vol. 29, no. 2–3, pp. 131–163, Nov. 1997, doi: 10.1023/A:1007465528199.
[32] J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink, and E. D. Jarvis, “Advances to Bayesian network inference for generating causal networks from observational biological data,” *Bioinformatics*, vol. 20, no. 18, pp. 3594–3603, Dec. 2004, doi: 10.1093/bioinformatics/bth448.
[33] B. Ellis and W. H. Wong, “Learning Causal Bayesian Network Structures from Experimental Data,” *J. Am. Stat. Assoc.*, vol. 103, no. 482, pp. 778–789, 2008.
[34] M. Scutari and J.-B. Denis, *Bayesian Networks: With Examples in R*. CRC Press, 2014.
[35] M. Scutari, “Bayesian Network Constraint-Based Structure Learning Algorithms: Parallel and Optimised Implementations in the bnlearn R Package,” *ArXiv14067648 Cs Stat*, Jun. 2014, Accessed: Mar. 26, 2018. [Online]. Available: http://arxiv.org/abs/1406.7648
[36] S. Meganck, P. Leray, and B. Manderick, “Learning Causal Bayesian Networks from Observations and Experiments: A Decision Theoretic Approach,” in *MDAI*, 2006. doi: 10.1007/11681960_8.
[37] D. Margaritis, “Learning Bayesian Network Model Structure from Data,” CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE, CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE, CMU-CS-03-153, May 2003. Accessed: Mar. 14, 2018. [Online]. Available: http://www.dtic.mil/docs/citations/ADA461103
[38] C. F. Aliferis, A. Statnikov, I. Tsamardinos, S. Mani, and X. D. Koutsoukos, “Local Causal and Markov Blanket Induction for Causal Discovery and Feature Selection for Classification Part I: Algorithms and Empirical Evaluation,” *J. Mach. Learn. Res.*, vol. 11, no. Jan, pp. 171–234, 2010.
[39] T. Verma and J. Pearl, *Equivalence and Synthese of Causal Models*. University of California (Los Angeles). Computer Science Department, 1991.
[40] R. Daly and Q. Shen, *Methods to Accelerate the Learning of Bayesian Network Structures*. 2007.
[41] T. Rauber and G. Rünger, *Parallel Programming: for Multicore and Cluster Systems*, 2nd ed. Berlin Heidelberg: Springer-Verlag, 2013. Accessed: Mar. 14, 2018. [Online]. Available: //www.springer.com/us/book/9783642378003
[42] M. Bartlett and J. Cussens, “Advances in Bayesian Network Learning using Integer Programming,” *ArXiv13096825 Cs*, Sep. 2013, Accessed: Mar. 14, 2018. [Online]. Available: http://arxiv.org/abs/1309.6825
[43] P. Spirtes, C. Glymour, S. N, and Richard, *Causation, Prediction, and Search*. Mit Press: Cambridge, 2000.
[44] J. Han and J. Pei, “Mining Frequent Patterns by Pattern-growth: Methodology and Implications,” *SIGKDD Explor Newsl*, vol. 2, no. 2, pp. 14–20, Dec. 2000, doi: 10.1145/380995.381002.
[45] I. Tsamardinos, C. Aliferis, A. Statnikov, and E. Statnikov, “Algorithms for Large Scale Markov Blanket Discovery,” in *In The 16th International FLAIRS Conference, St*, AAAI Press, 2003, pp. 376–380.
[46] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2009.
[47] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, “The max-min hill-climbing Bayesian network structure learning algorithm,” *Mach. Learn.*, vol. 65, no. 1, pp. 31–78, Oct. 2006, doi: 10.1007/s10994-006-6889-7.
[48] H. Shimodaira, “Improving predictive inference under covariate shift by weighting the log-likelihood function,” *J. Stat. Plan. Inference*, vol. 90, no. 2, pp. 227–244, Oct. 2000, doi: 10.1016/S0378-3758(00)00115-4.
[49] J. Kim and K.-Y. Chung, “Ontology-based healthcare context information model to implement ubiquitous environment,” *Multimed. Tools Appl.*, vol. 71, no. 2, pp. 873–888, Jul. 2014, doi: 10.1007/s11042-011-0919-6.
[50] D. G. MacArthur *et al.*, “Guidelines for investigating causality of sequence variants in human disease,” *Nature*, vol. 508, no. 7497, pp. 469–476, Apr. 2014, doi: 10.1038/nature13127.
[51] F. Alemi, M. Zargoush, and J. Vang, “Using observed sequence to orient causal networks,” *Health Care Manag. Sci.*, Jul. 2016, doi: 10.1007/s10729-016-9373-3.
[52] O. A. Arah, “The role of causal reasoning in understanding Simpson’s paradox, Lord’s paradox, and the suppression effect: covariate selection in the analysis of observational studies,” *Emerg. Themes Epidemiol.*, vol. 5, p. 5, Feb. 2008, doi: 10.1186/1742-7622-5-5.
[53] Y.-B. He and Z. Geng, “Active Learning of Causal Networks with Intervention Experiments and Optimal Designs,” 2008.
[54] S. Triantafillou, I. Tsamardinos, and I. G. Tollis, “Learning Causal Structure from Overlapping Variable Sets,” in *AISTATS*, 2010.
[55] L. J. Rips, “Causation From Perception,” *Perspect. Psychol. Sci.*, vol. 6, no. 1, pp. 77–97, 2011.
[56] L. A. Goodman and W. H. Kruskal, “Measures of Association for Cross Classifications*,” *J. Am. Stat. Assoc.*, vol. 49, no. 268, pp. 732–764, 1954, doi: 10.1080/01621459.1954.10501231.
[57] C. R. Levy *et al.*, “Sequence of Functional Loss and Recovery in Nursing Homes,” *The Gerontologist*, vol. 56, no. 1, pp. 52–61, 2016, doi: 10.1093/geront/gnv099.

[58] M. Ben Messaoud, P. Leray, and N. Ben Amor, “SemCaDo: A serendipitous strategy for causal discovery and ontology evolution,” *Knowl.-Based Syst.*, vol. 76, pp. 79–95, Mar. 2015, doi: 10.1016/j.knosys.2014.12.006.
[59] M. B. Ishak, P. Leray, and N. B. Amor, “Ontology-based generation of object oriented bayesian networks,” presented at the BMAW 2011, 2011, p. 9. Accessed: Apr. 27, 2023. [Online]. Available: https://hal.science/hal-00644992
[60] M. Munch, J. Dibie-Barthelemy, P.-H. Wuillemin, and C. Manfredotti, “Towards Interactive Causal Relation Discovery Driven by an Ontology,” presented at the FLAIRS 32, May 2019. Accessed: Apr. 27, 2023. [Online]. Available: https://hal.science/hal-02184398
[61] M. Scutari, “Learning Bayesian Networks with the bnlearn R Package,” *ArXiv09083817 Stat*, Aug. 2009, Accessed: May 14, 2017. [Online]. Available: http://arxiv.org/abs/0908.3817
[62] A. P. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997, doi: 10.1016/S0031-3203(96)00142-2.
[63] H. Hu, A. Elkus, and L. Kerschberg, “A Personal Health Recommender System incorporating personal health records, modular ontologies, and crowd-sourced data,” in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Aug. 2016, pp. 1027–1033. doi: 10.1109/ASONAM.2016.7752367.
[64] H. Hu and L. Kerschberg, “Standardizing the Crowdsourcing of Healthcare Data Using Modular Ontologies,” in *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, Jul. 2017, pp. 107–112. doi: 10.1109/COMPSAC.2017.220.
[65] H. Hu, A. H. Elrafey, and L. Kerschberg, “Using Modular Ontologies to Capture Causal Knowledge Contained in Bayesian Networks,” in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, in ASONAM ’17. New York, NY, USA: ACM, 2017, pp. 904–907. doi: 10.1145/3110025.3110136.
[66] H. Hu and L. Kerschberg, “Evolving Medical Ontologies Based on Causal Inference,” in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Aug. 2018, pp. 954–957. doi: 10.1109/ASONAM.2018.8508259.
[67] H. Hu and L. Kerschberg, “Improving Causal Bayesian Networks Using Expertise in Authoritative Medical Ontologies,” *ACM Trans. Comput. Healthc.*, vol. 4, no. 4, p. 20:1-20:32, Oct. 2023, doi: 10.1145/3604561.
[68] B. Andrea and T. Franco, “Mining Bayesian networks out of ontologies,” *J. Intell. Inf. Syst.*, vol. 38, no. 2, pp. 507–532, Apr. 2012, doi: 10.1007/s10844-011-0165-4.
[69] S. Fenz, A. M. Tjoa, and M. Hudec, “Ontology-Based Generation of Bayesian Networks,” in *2009 International Conference on Complex, Intelligent and Software Intensive Systems*, Mar. 2009, pp. 712–717. doi: 10.1109/CISIS.2009.33.
[70] S. Fenz, “An ontology-based approach for constructing Bayesian networks,” *Data Knowl. Eng.*, vol. 73, pp. 73–88, Mar. 2012, doi: 10.1016/j.datak.2011.12.001.
[71] R. Pan, Z. Ding, Y. Yu, and Y. Peng, “A Bayesian Network Approach to Ontology Mapping,” in *The Semantic Web – ISWC 2005*, Y. Gil, E. Motta, V. R. Benjamins, and M. A. Musen, Eds., in Lecture Notes in Computer Science. Springer Berlin Heidelberg, Nov. 2005, pp. 563–577. doi: 10.1007/11574620_41.
[72] S. Matsumoto, R. N. Carvalho, P. C. Costa, K. B. Laskey, L. L. Dos Santos, and M. Ladeira, *There’s No More Need to be a Night OWL: on the PR-OWL for a MEBN Tool Before Nightfall*. iConcept Press, 2011. Accessed: Mar. 24, 2017. [Online]. Available: https://pdfs.semanticscholar.org/bebd/6ab79ecacf96dbd58261fe2e6362bdfdfecb.pdf
[73] Y. Yang and J. Calmet, “OntoBayes: An Ontology-Driven Uncertainty Model,” in *Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce Vol-1 (CIMCA-IAWTIC’06) - Volume 01*, in CIMCA ’05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 457–463. Accessed: Oct. 29, 2016. [Online]. Available: http://dl.acm.org/citation.cfm?id=1134823.1135162
[74] K. P. Murphy, “Active Learning of Causal Bayes Net Structure,” 2001.
[75] S. Meganck, “Towards an Integral Approach for Modeling Causality,” 2008.
[76] Z. Zhao *et al.*, “Recommender Systems in the Era of Large Language Models (LLMs),” *IEEE Trans. Knowl. Data Eng.*, pp. 1–20, 2024, doi: 10.1109/TKDE.2024.3392335.
[77] L. A. Cox Jr., “Pushing Back on AI: A Dialogue with ChatGPT on Causal Inference in Epidemiology,” in *AI-ML for Decision and Risk Analysis: Challenges and Opportunities for Normative Decision Theory*, L. A. Cox Jr., Ed., in International Series in Operations Research & Management Science. , Cham: Springer International Publishing, 2023, pp. 407–423. doi: 10.1007/978-3-031-32013-2_13.
[78] M. Gleichmann, V. W. Chow, and M. P. Mattson, “Homeostatic Disinhibition in the Aging Brain and Alzheimer’s Disease,” *J. Alzheimers Dis. JAD*, vol. 24, no. 1, pp. 15–24, 2011, doi: 10.3233/JAD-2010-101674.
[79] S. Nash, J. D. Henry, S. Mcdonald, I. Martin, H. Brodaty, and M.-A. Peek-O’leary, “Cognitive disinhibition and socioemotional functioning in Alzheimer’s disease,” *J. Int. Neuropsychol. Soc.*, vol. 13, no. 6, pp. 1060–1064, Nov. 2007, doi: 10.1017/S1355617707071184.
[80] F. G. de M. Coelho *et al.*, “Gait and risk of falls associated with frontal cognitive functions at different stages of Alzheimer’s disease,” *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.*, vol. 19, no. 5, pp. 644–656, Sep. 2012, doi: 10.1080/13825585.2012.661398.
[81] W. E. Schell, V. S. Mar, and C. P. Da Silva, “Correlation of falls in patients with Amyotrophic Lateral Sclerosis with objective measures of balance, strength, and spasticity,” *Neurorehabilitation*, vol. 44, no. 1, pp. 85–93, doi: 10.3233/NRE-182531.
[82] O. L. Lopez, J. T. Becker, R. P. Brenner, J. Rosen, O. I. Bajulaiye, and C. F. Reynolds, “Alzheimer’s disease with delusions and hallucinations: Neuropsychological and electroencephalographic correlates,” *Neurology*, vol. 41, no. 6, pp. 906–906, Jun. 1991, doi: 10.1212/WNL.41.6.906.
[83] N. Scarmeas *et al.*, “Delusions and Hallucinations Are Associated With Worse Outcome in Alzheimer Disease,” *Arch. Neurol.*, vol. 62, no. 10, Oct. 2005, doi: 10.1001/archneur.62.10.1601.
[84] R. S. Wilson, “Hallucinations, delusions, and cognitive decline in Alzheimer’s disease,” *J. Neurol. Neurosurg. Psychiatry*, vol.

69, no. 2, pp. 172–177, Aug. 2000, doi: 10.1136/jnnp.69.2.172.

[85] G. Coughlan, J. Laczó, J. Hort, A.-M. Minihane, and M. Hornberger, “Spatial navigation deficits - overlooked cognitive marker for preclinical Alzheimer disease?,” *Nat. Rev. Neurol.*, vol. 14, no. 8, pp. 496–506, Aug. 2018, doi: 10.1038/s41582-018-0031-x.

[86] V. Puthusseryppady, L. Emrich-Mills, E. Lowry, M. Patel, and M. Hornberger, “Spatial Disorientation in Alzheimer’s Disease: The Missing Path From Virtual Reality to Real World,” *Front. Aging Neurosci.*, vol. 12, 2020, doi: 10.3389/fnagi.2020.550514.

[87] A. M. Monacelli, L. A. Cushman, V. Kavcic, and C. J. Duffy, “Spatial disorientation in Alzheimer’s disease: The remembrance of things passed,” *Neurology*, vol. 61, no. 11, pp. 1491–1497, Dec. 2003, doi: 10.1212/WNL.61.11.1491.

[88] H. L. O’Brien, S. J. Tetewsky, L. M. Avery, L. A. Cushman, W. Makous, and C. J. Duffy, “Visual Mechanisms of Spatial Disorientation in Alzheimer’s Disease,” *Cereb. Cortex*, vol. 11, no. 11, pp. 1083–1092, Nov. 2001, doi: 10.1093/cercor/11.11.1083.

[89] M.-C. Pai and W. J. Jacobs, “Topographical disorientation in community-residing patients with Alzheimer’s disease,” *Int. J. Geriatr. Psychiatry*, vol. 19, no. 3, pp. 250–255, 2004, doi: https://doi.org/10.1002/gps.1081.

[90] A. R. Powers, M. Kelley, and P. R. Corlett, “Hallucinations as top-down effects on perception,” *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, vol. 1, no. 5, pp. 393–400, Sep. 2016, doi: 10.1016/j.bpsc.2016.04.003.

[91] G. E. Berrios, “Disorientation states and psychiatry,” *Compr. Psychiatry*, vol. 23, no. 5, pp. 479–491, Sep. 1982, doi: 10.1016/0010-440X(82)90161-4.

[92] M. Mendez, R. Tomsak, and B. Remler, “Disorders of the visual system in Alzheimer’s Disease,” *J. Clin. Neuroophthalmol.*, vol. 10, pp. 62–9, Apr. 1990.

[93] O. L. Lopez *et al.*, “Psychiatric Symptoms Vary With the Severity of Dementia in Probable Alzheimer’s Disease,” *J. Neuropsychiatry Clin. Neurosci.*, vol. 15, no. 3, pp. 346–353, Aug. 2003, doi: 10.1176/jnp.15.3.346.

[94] H.-C. Cheng, C. M. Ulane, and R. E. Burke, “Clinical Progression in Parkinson’s Disease and the Neurobiology of Axons,” *Ann. Neurol.*, vol. 67, no. 6, pp. 715–725, Jun. 2010, doi: 10.1002/ana.21995.

[95] E. Kim, “Agitation, aggression, and disinhibition syndromes after traumatic brain injury,” *NeuroRehabilitation*, vol. 17, no. 4, pp. 297–310, Jan. 2002, doi: 10.3233/NRE-2002-17404.

[96] T. H. Reekes, C. I. Higginson, C. R. Ledbetter, N. Sathivadivel, R. M. Zweig, and E. A. Disbrow, “Sex specific cognitive differences in Parkinson disease,” *Npj Park. Dis.*, vol. 6, no. 1, Art. no. 1, Apr. 2020, doi: 10.1038/s41531-020-0109-1.

[97] B. Yew, S. Alladi, M. Shailaja, J. R. Hodges, and M. Hornberger, “Lost and forgotten? Orientation versus memory in Alzheimer’s disease and frontotemporal dementia,” *J. Alzheimers Dis. JAD*, vol. 33, no. 2, pp. 473–481, 2013, doi: 10.3233/JAD-2012-120769.

[98] F. Bunn *et al.*, “Comorbidity and dementia: a scoping review of the literature,” *BMC Med.*, vol. 12, Oct. 2014, doi: 10.1186/s12916-014-0192-4.