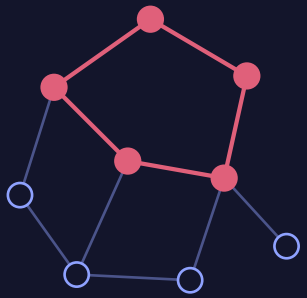
IT PHD SEMINAR · ILLINOIS TECH · NOVEMBER 6, 2026

Asking Machines Why



Explanation and reasoning in graphs, memory, and language models

Arman Behnam

PhD candidate, Computer Science · Illinois Institute of Technology · Advisor: Binghui Wang

OPENING

Your card is declined

BANK ALERT · 7:42 PM

Payment declined

$42.10 · grocery store · flagged

Why?

- New city?
- Unusual amount?
- Odd hour?

OPENING

Two questions

QUESTION 1

Is it right?

Check against outcomes

QUESTION 2

Why?

Ask: what if?

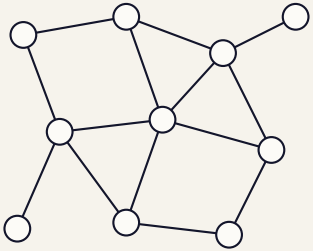
TODAY

01 Graphs **02 Memory** **03 Language models** **04 The thread** **05 My answers**

01
Graphs

Graphs | Memory | Language models | The thread | My answers

A graph: things and their links



NODES = THINGS · EDGES = LINKS

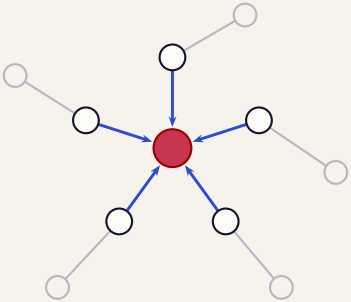
Friends
people · who knows whom

Molecules
atoms · bonds

Roads
junctions · streets

Payments
accounts · transfers

A GNN asks its neighbors



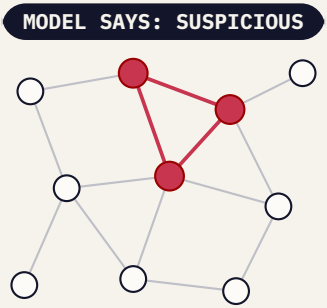
01 Own info

02 Neighbor messages

03 Repeat → predict

GNN: graph neural network

An explanation: what mattered?

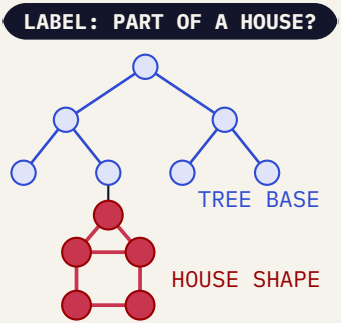


01 Input: one prediction

02 Output: a small piece

03 **Rule: most predictive**

The explainer picks the tree

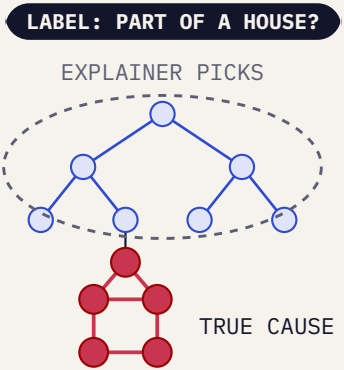


01 Houses sit on trees

02 Other shapes, other bases

03 Tree predicts the label

The explainer picks the tree



01 Houses sit on trees

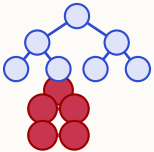
02 Other shapes, other bases

03 Tree predicts the label

Most predictive → the tree

Causal questions ask: what if?

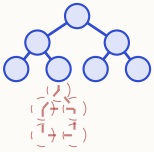
SEEING



Together

Tree predicts too

DOING



Remove house

Label changes

What if? needs a cause-and-effect model

Where do we get the cause-and-effect model?

Hold this question · Next: Memory

02
Memory

Graphs | **Memory** | Language models | The thread | My answers

Assistants keep notes

MONDAY

I’m allergic to peanuts.

FRIDAY, A NEW CONVERSATION

Suggest a quick dinner.

MODEL ALONE

Try spicy peanut noodles!

WITH MEMORY

Here is a peanut-free stir-fry.

Where memory lives

01

Inside the model

From training · fixed

02

In the chat

On screen · temporary

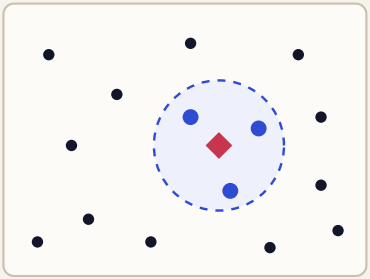
03 · THIS TALK

In a store

Saved notes · looked up

Finding a note = retrieval

Vector search: closest notes win



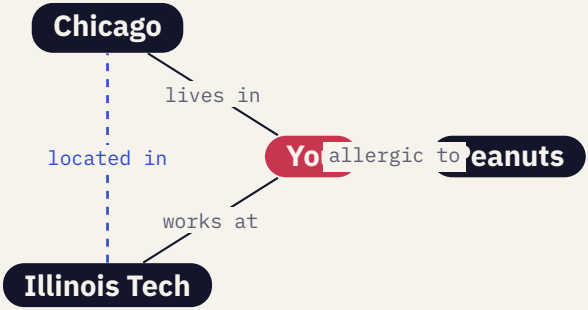
◆ your question ● top-3 retrieved • other notes

01 Note → point

02 Question → point

03 Closest *k* → retrieved

Graph memory: follow the links



01
Facts as links

02
Hop between facts

03
Hybrid: closest + links

GRAPHS, AGAIN

Which note helped?

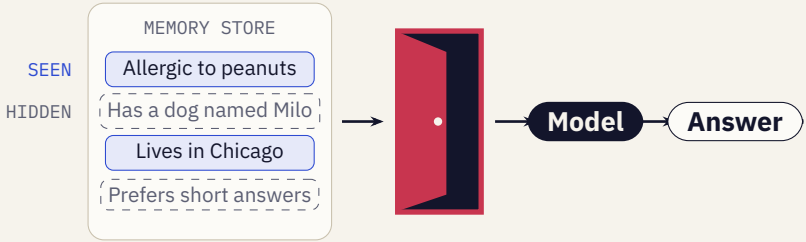
- Allergic to peanuts ?
- Lives in Chicago ?
- Prefers short answers ?
- Has a dog named Milo ?
- Asked about flights in May ?

01 Thousands of notes

02 Keep · delete · rank

Value: how much the note improves the answer

Retrieval is a door



Only retrieved notes show their value

54% · 67%

OF NEEDED NOTES HIDDEN · 2 MEMORY BENCHMARKS

How can we measure the value of a hidden note?

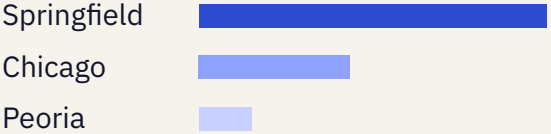
Hold this one too · Next: Language models

03
Language models

Graphs | Memory | **Language models** | The thread | My answers

Language models: the next word

The capital of Illinois is ___



ILLUSTRATIVE SCORES

01 Trained on text

02 One word at a time

Ask it why

Why was my card declined?

MODEL

Because the amount was unusual for you.

HIDDEN CUE: ODD HOUR

01 Fluent reason

02 Stated reason may differ from the real one

Turpin et al., NeurIPS 2023

Test reasoning directly

Reasoning, or recalling?

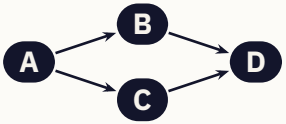
FAMILIAR NAMES

Does smoking cause cancer?

Recall is enough

MADE-UP NAMES

Does A cause D?



Reasoning needed

Zečević et al., TMLR 2023 · Jin et al., NeurIPS 2023

A hidden cause hides the answer



01 Seen: coffee, grades

02 Hidden: motivation

03 Two stories, same data

A hidden cause hides the answer



01 Seen: coffee, grades

02 Hidden: motivation

03 Two stories, same data

Honest answer: undetermined

Does the model know when the honest answer is “undetermined”?

Next: the thread

04
The thread

Graphs | Memory | Language models | **The thread** | My answers

One gap, three systems

SYSTEM	WHO ASKS	THE WHY	THE GAP
Graphs	We	Which part caused it?	House and tree together
Memory	We	Which note helped?	Hidden notes
Language models	**The model**	Answer or “undetermined”?	Hidden cause

Data shows what goes together. We ask what if.

First: can we answer it?

1 Can the data + what we know answer it?

2 Then: how accurate is the estimate?

Step 1 = identification

Three ways forward

01

Add knowledge

Build a causal model

02

Change the data

Collect it differently

03

Say “undetermined”

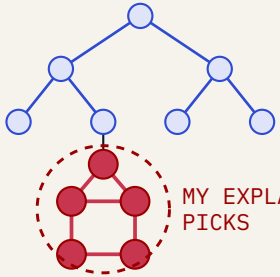
Honest answer

My work: one way per system

05
My answers

Graphs | Memory | Language models | The thread | My answers

Graphs: build the causal model



01 Pick a node

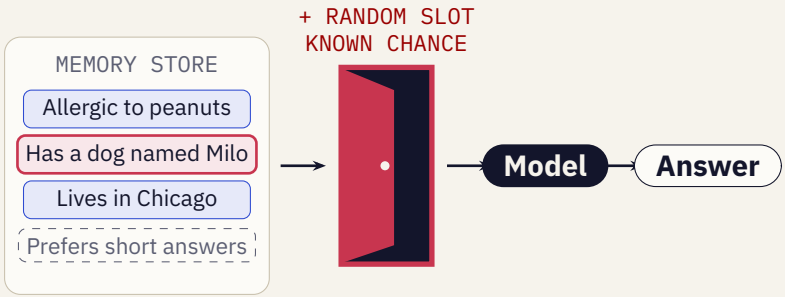
02 Message paths → causal structure

03 Train a causal model → what if?

99.4%
FINDS THE EXACT HOUSE

≤4%
OTHER EXPLAINERS

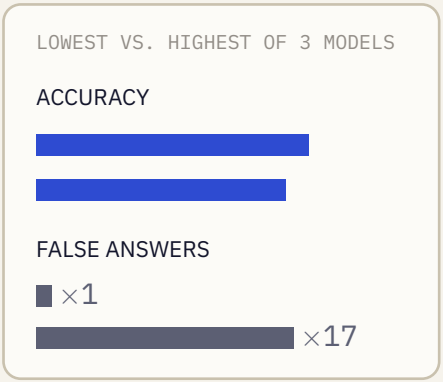
Memory: open the door on purpose



Every note gets tested

COST: A FEW CONTEXT SLOTS

Similar accuracy, 17× apart



01 3 top chat models

02 Proven answer key

03 Diagrams: 10–50 nodes

Answer: depends on the model

The whole journey

GRAPHS · TODAY

Build it

Construct the causal model

IMAGES

Assume it

Know only the direction

STRUCTURE

Learn it

From many environments

MEMORY · TODAY

Collect it

Change what is retrieved

LANGUAGE · TODAY

Test it

Can models say “undetermined”?

First: can we answer it?

Open questions

1. **Causal explanations for real fraud graphs?**
2. **Design systems that can answer why?**
3. **Models that know when to say “undetermined”?**

THANK YOU

Join me at my defense

PHD DISSERTATION DEFENSE

Thursday, November 19, 2026

12:30–1:45 PM CT · SB 233

Zoom

Questions?