Massachusetts General Hospital
Founding Member, Mass General Brigham

HARVARD
MEDICAL SCHOOL

A Multi-dimensional index for evaluating Systems Thinking Skills from Textual Data
&
Causal Loop Diagram Generation using Generative AI

Georgia Ning-Yuan Liu, Ph.D.

Postdoctoral Research Fellow
Massachusetts General Hospital Institute for Technology Assessment
Harvard Medical School

gliu26@mgh.harvard.edu

www.linkedin.com/in/ngeorgialiu

https://georgia-max.github.io/

09/27/2024 @Purdue system thinkers club

Research Journey

COLLEGE OF ENGINEERING
GRADO DEPARTMENT OF
INDUSTRIAL AND SYSTEMS ENGINEERING
VIRGINIA TECH.

SPL | SYSTEM PERFORMANCE LAB

Dr. Konstantinos Triantis

Empowering the understanding of the **management of complex socio-technical systems** through Data Science and Systems Thinking

HARVARD
MEDICAL SCHOOL

MGH Institute for Technology Assessment

MJ LAB

Data and Systems Science in
Public Health

Dr. Mohammad Jalali

Understanding **complex public health problems** to inform decision and policy-making through Data Science and Simulation Modelling

Failed to Understand Complex Systems leads to Disaster

Transportation

Finance

nancial crisis
Crash of 200
Disbelief, and a punt

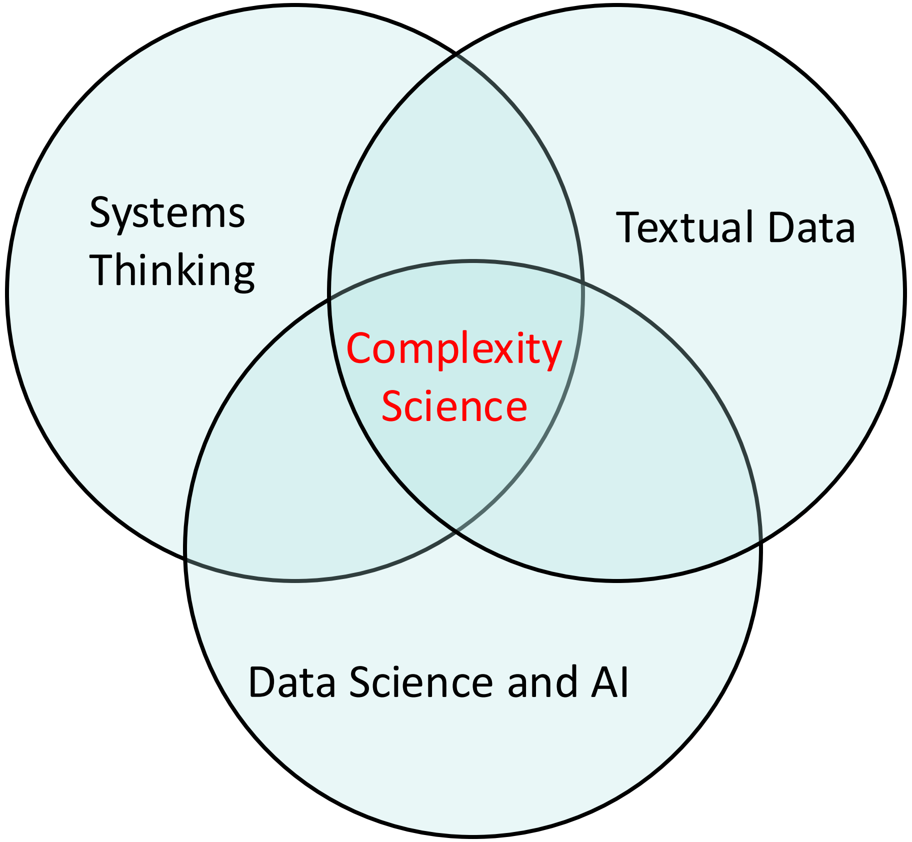
Environment

Public Health

Research Focus: Understanding Complexity Through …

Transportation

Environment

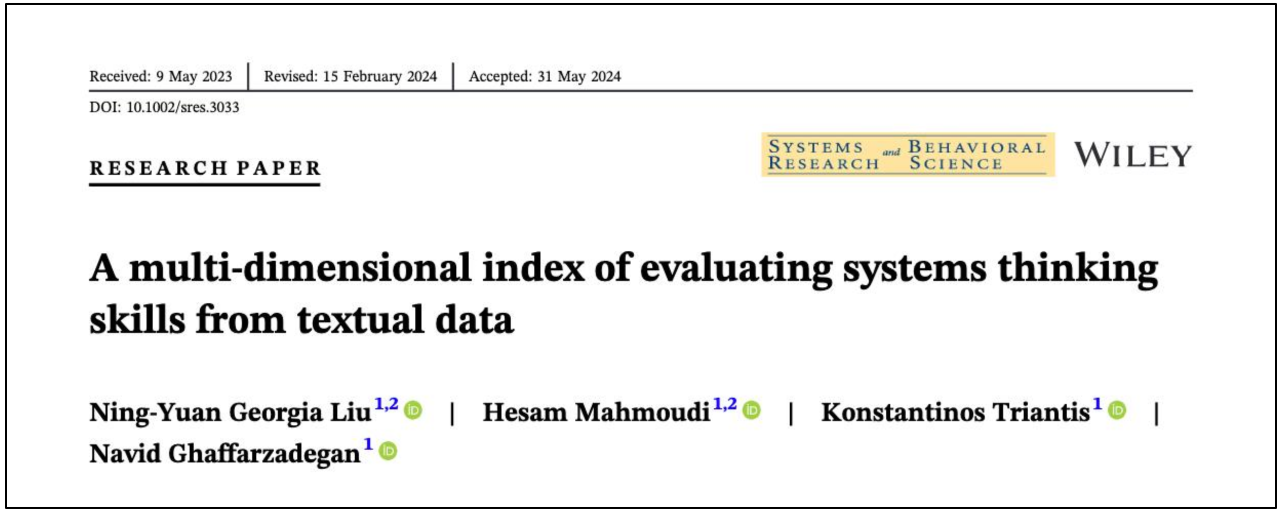


Finance

ancial crisis
Crash of 200
Disbelief, and a punt

Public Health

Today’s Topic



Received: 9 May 2023 | Revised: 15 February 2024 | Accepted: 31 May 2024

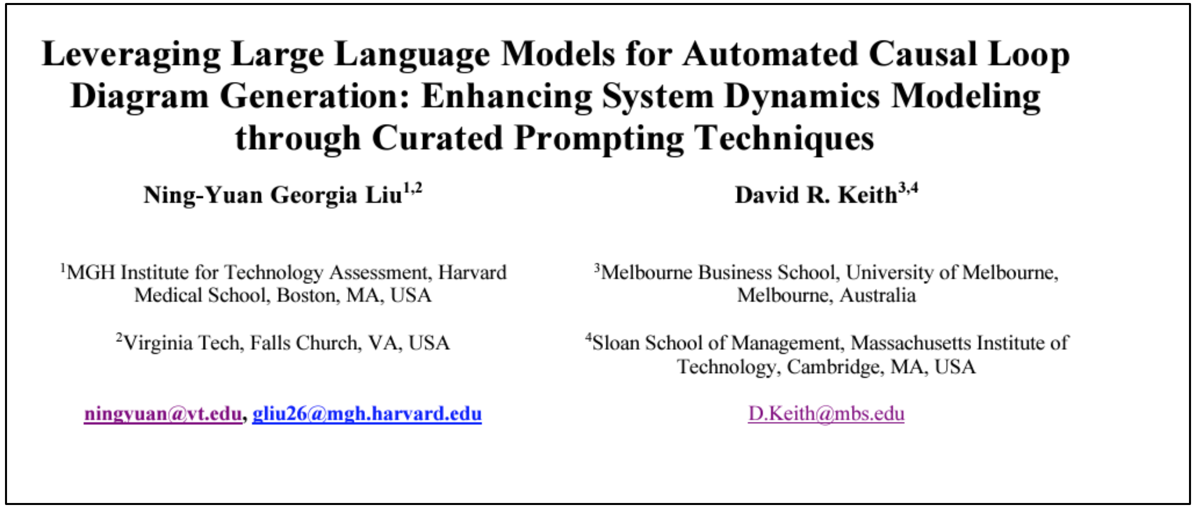
DOI: 10.1002/sres.3033

RESEARCH PAPER

SYSTEMS and BEHAVIORAL RESEARCH SCIENCE WILEY

A multi-dimensional index of evaluating systems thinking skills from textual data

Ning-Yuan Georgia Liu[1,2] | Hesam Mahmoudi[1,2] | Konstantinos Triantis[1] | Navid Ghaffarzadegan[1]



Leveraging Large Language Models for Automated Causal Loop Diagram Generation: Enhancing System Dynamics Modeling through Curated Prompting Techniques

Ning-Yuan Georgia Liu[1,2]

[1]MGH Institute for Technology Assessment, Harvard Medical School, Boston, MA, USA

[2]Virginia Tech, Falls Church, VA, USA

ningyuan@vt.edu, gliu26@mgh.harvard.edu

David R. Keith[3,4]

[3]Melbourne Business School, University of Melbourne, Melbourne, Australia

[4]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA

D.Keith@mbs.edu

Dr. H. Mahmoudi

Dr. N. Ghaffarzadegan

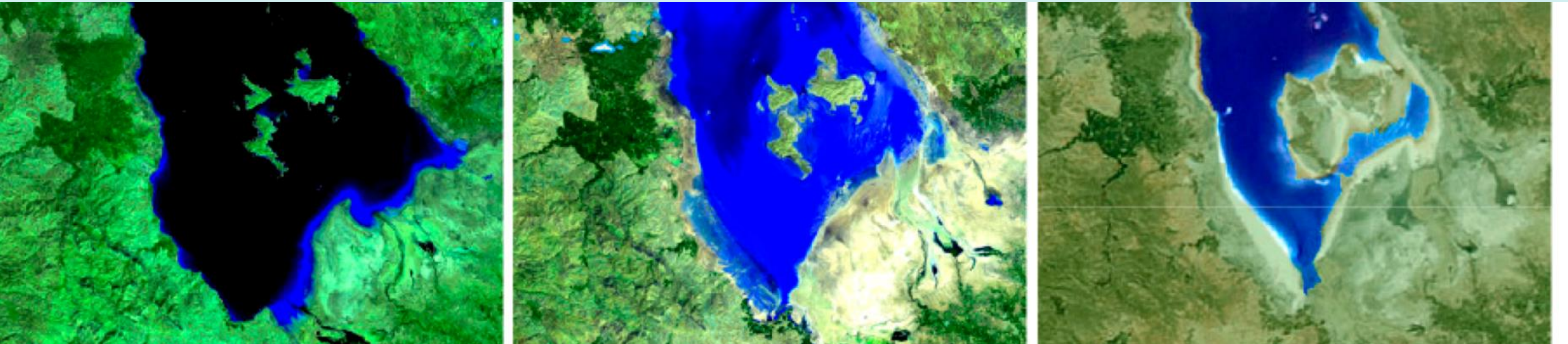
Dr. K. Triantis

Dr. D. Keith

NSF Grant #1824594: Cognitive Barriers to Understanding Complexity in Human-Technical Systems: Evidence from Engineering Students and Practitioners

Paper I

A Multi-dimensional index for evaluating Systems Thinking Skills from Textual Data



Outline

Introduction

- System Thinking (ST)
- Challenges of Multidimensionality in ST
- Research Questions

Data & Methods

- System Thinking Measures
- Multi-dimensional Index of Systems Thinking (MIST)
- Benchmarking Process

Results

- Comprehensive ST Skills Performance
- The Impact of Contextual Variables on MIST scores

Conclusions

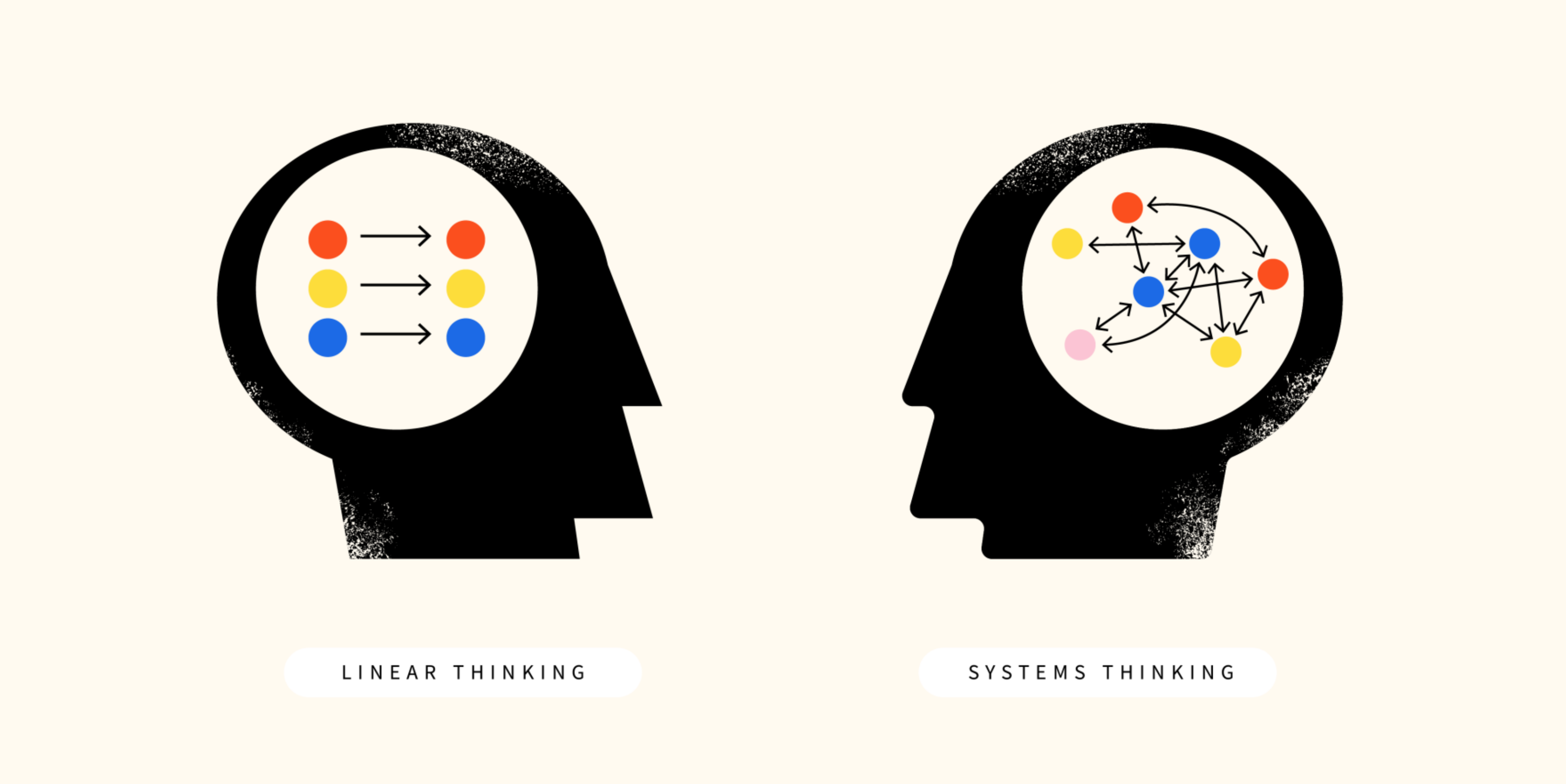
- Lesson Learnt
- Limitations and Future Research

Introduction

- System Thinking (ST)
- Challenges of Multidimensionality in ST
- Research Questions

Systems thinking:

a way of making sense of the **complexity** of the world by looking at it in terms of wholes and relationships rather than by splitting it down into its parts…



System Thinking has a Multi-Dimensional Characteristic

Approach	Richmond (1994)	STH: Assaraf and Orion (2005)	STC: Stave and Hopper (2007)
Level 1	*Specify Problems:* ➢ **Forest Thinking** ➢ System as Cause Thinking ➢ Dynamic Thinking	*System components:* ➢ The ability to identify the **components** of a system and processes within the system	*Basic:* ➢ Recognizing interconnections ➢ Identifying **feedback** ➢ Understanding **dynamic behaviors**
Level 2	*Construct Model:* ➢ Quantitative thinking ➢ **Closed Loop Thinking** ➢ Operational Thinking	*Synthesis of system components:* ➢ identify relationships among system’s components ➢ organize the systems’ components and processes ➢ identify **dynamic relationships**	*Intermediate:* ➢ Differentiations types of flow and variables ➢ Using conceptual models
Level 3	*Test Model:* ➢ Scientific thinking	*Implementation:* ➢ Understanding the hidden dimensions of the system ➢ The ability to understand the **cyclic nature** of systems ➢ Thinking temporally: retrospection and prediction	*Advanced:* ➢ Creating simulation models ➢ Testing policies

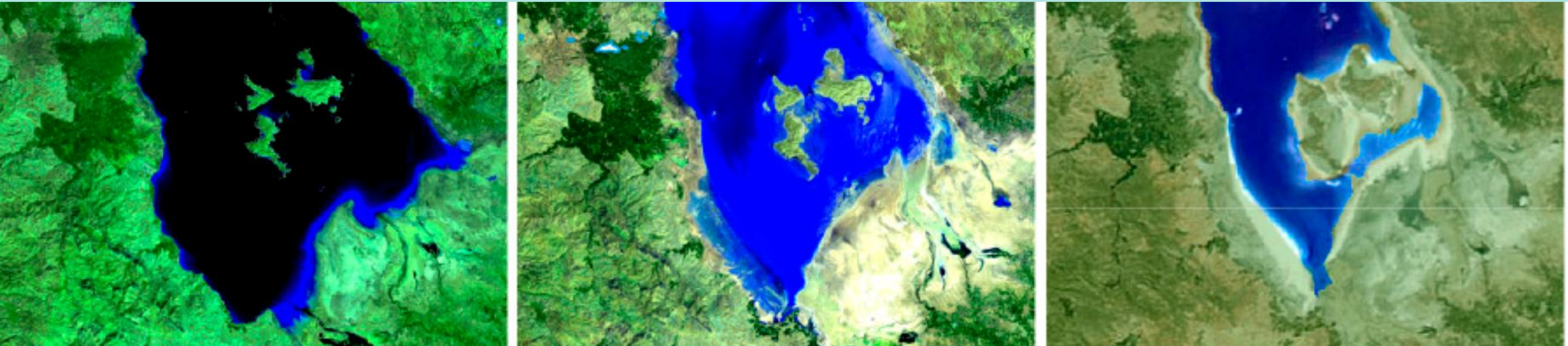
Research Questions

1. How do we measure one’s level of System Thinking (ST) ?

2. How do we compute a multi-dimensional ST index where ST skill characteristics are considered concurrently?

Data & Methods

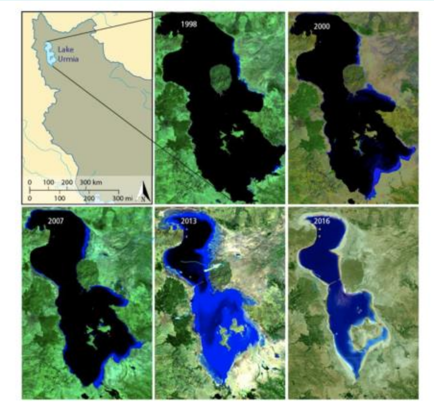
- The Lake Urmia Vignette
- Mental Map Measures
- Multidimensional Comprehension Index of Systems Thinking (MCIST) Framework



Translating Textual Data to Mental Maps

The Lake Urmia Vignette

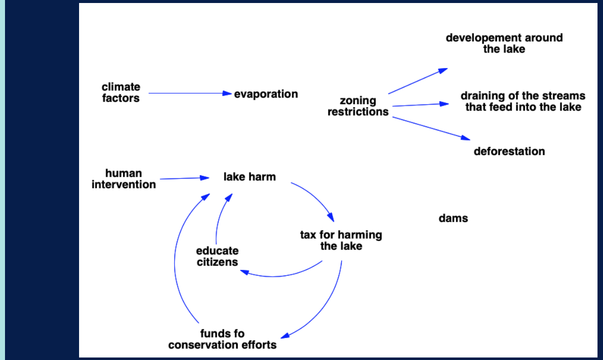
Q: Why is the lake shrinking?



Participant Response in Text format

I think the problem was that, as technology developed, it led to destruction of nature and caused excessive evaporation of water. Cities stole trees and water for their development and didn't really think about Lake Urmia's condition. Now everyone can obviously see what has changed. Plant more trees and create an eco-friendly environment that will stop water evaporation and bring back animal inhabitants.

ST measures



Color Coding

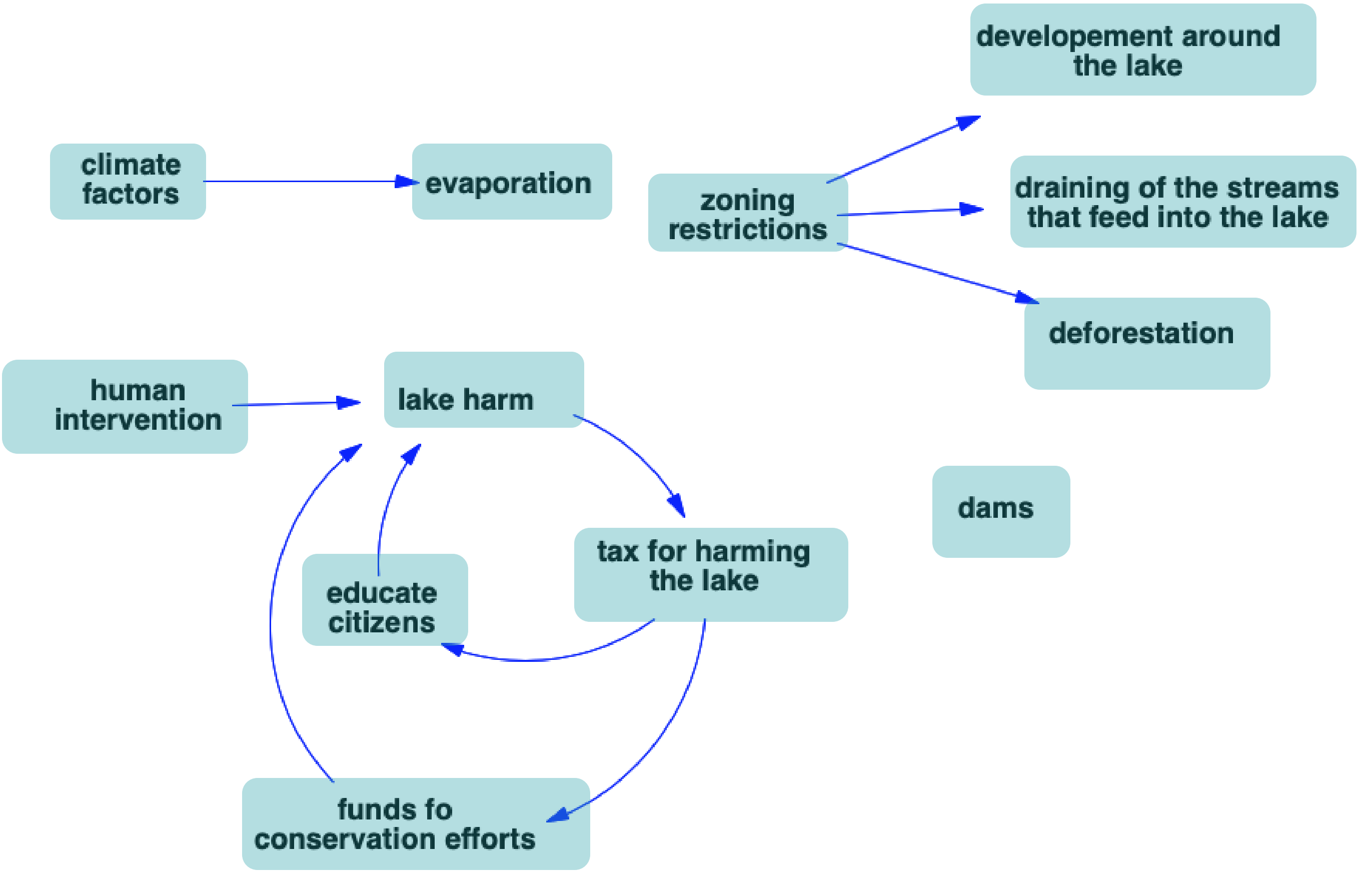
I think the problem was that, as technology developed, it led to destruction of nature and caused excessive evaporation of water. Cities stole trees and water for their development and didn't really think about Lake Urmia's condition. Now everyone can obviously see what has changed. Plant more trees and create an eco-friendly environment that will stop water evaporation and bring back animal inhabitants.

Word-arrow relation

- technology → (↑) destruction of nature & (↓) evaporation of water (2 causal effects)
- development of cities → (↓) trees & (↓) water (2 causal effects)
- trees → (↑) eco-friendly environment → (↓) water evaporation → (↑) animal inhabitants (3 causal effects)

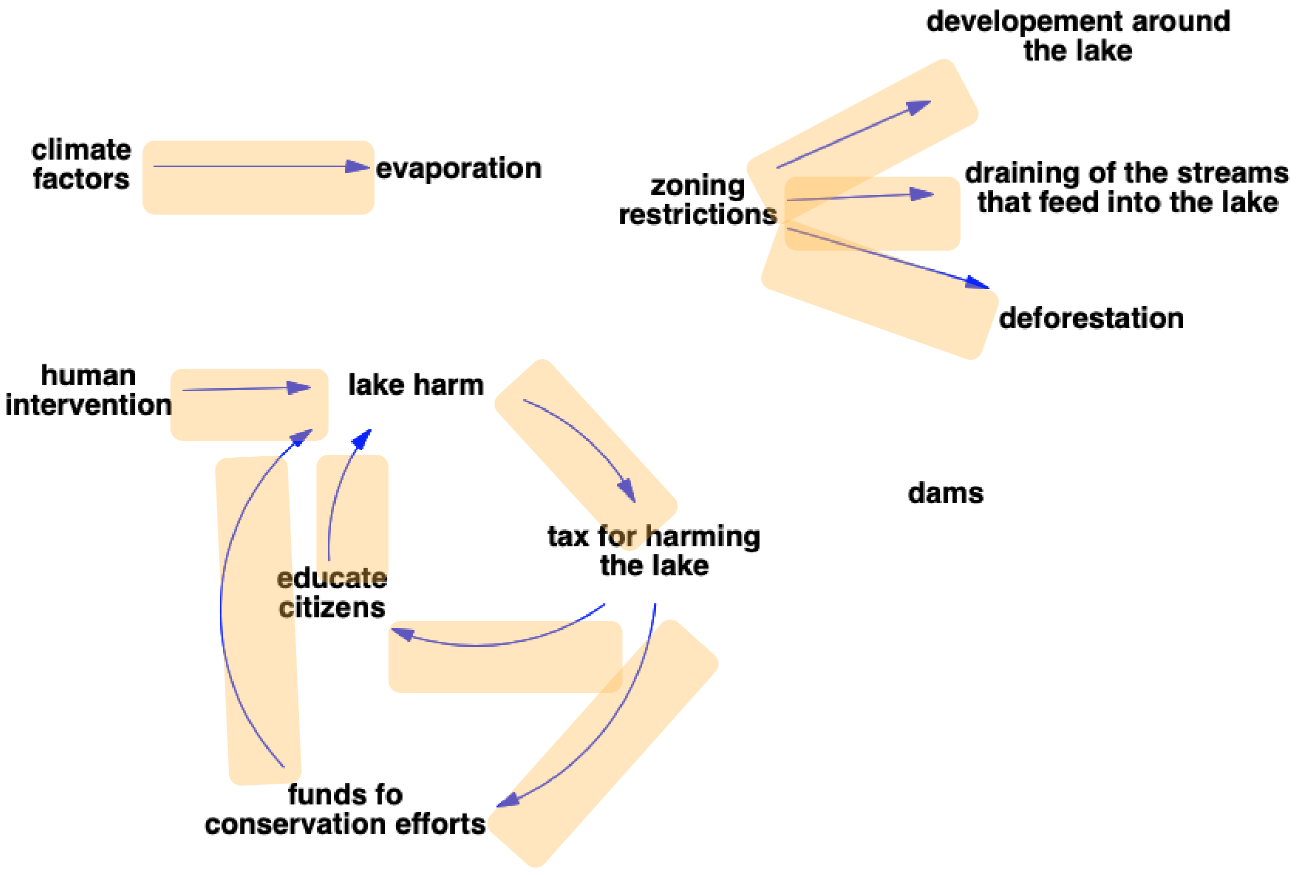
Davis et al. (2020) The Lake Urmia vignette: A tool to assess understanding of complexity in socio-environmental systems. System Dynamics Review, 36(2), 191–222.

1. ST Measures – Number of Variables



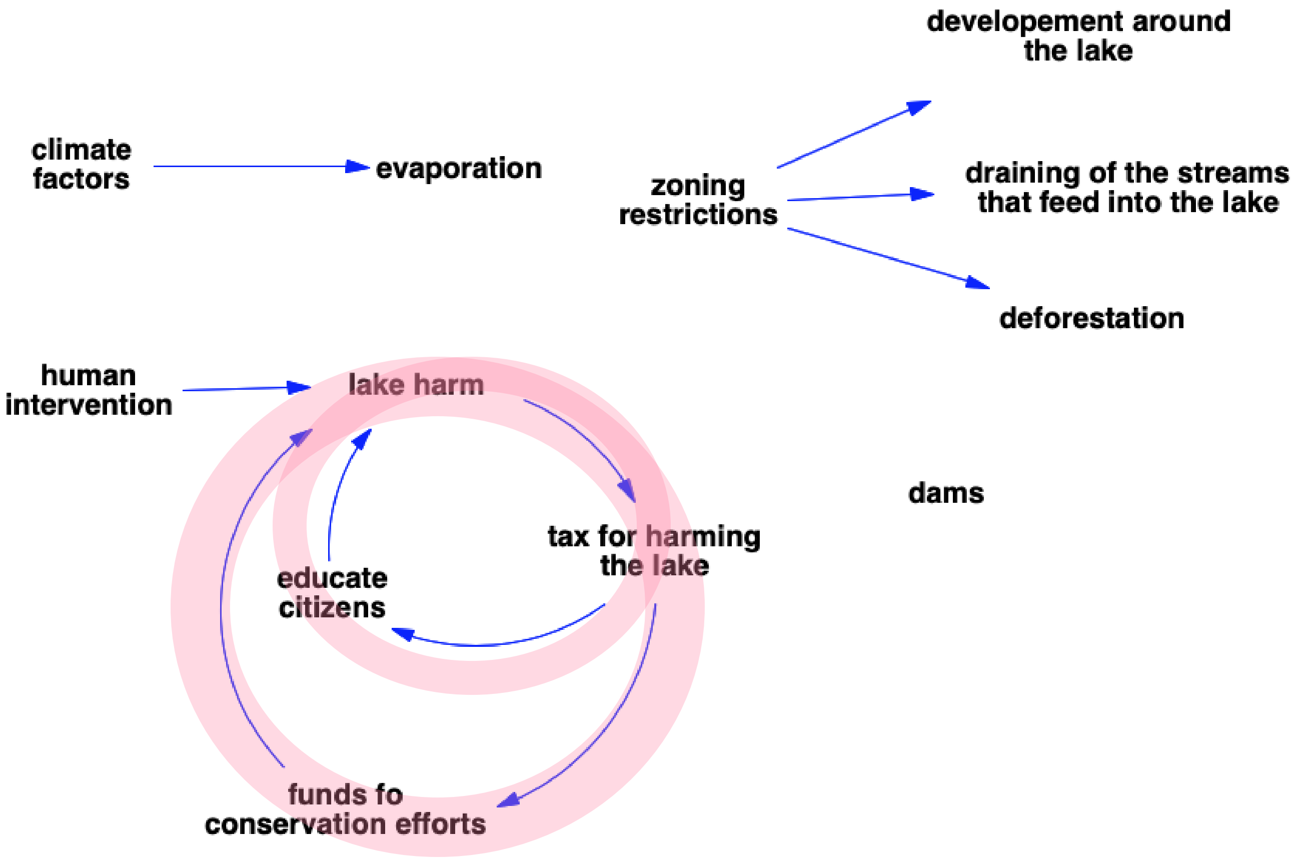
Measure	#	Type of ST
Variables	11	✓ Detailed complexity (Richardson, 1994) ✓ Dynamic thinking (Stave & Hopper, 2007)
Causal links		✓ Interconnectivity (Dorani et al., 2015) ✓ Cause-effect thinking (Stave & Hopper, 2007)
Closed loops		✓ System-as-cause thinking (Dorani et al., 2015) ✓ Closed-loop thinking (Richardson, 1994) ✓ Identifying feedback (Stave & Hopper, 2007)
Middle Nodes		✓ Operational thinking (Haque et al., 2023)
Connectivity		✓ Cyclomatic complexity (Naugle et al., 2021) ✓ Link density (Plate, 2010)

2. ST Measures – Number of Casual Links



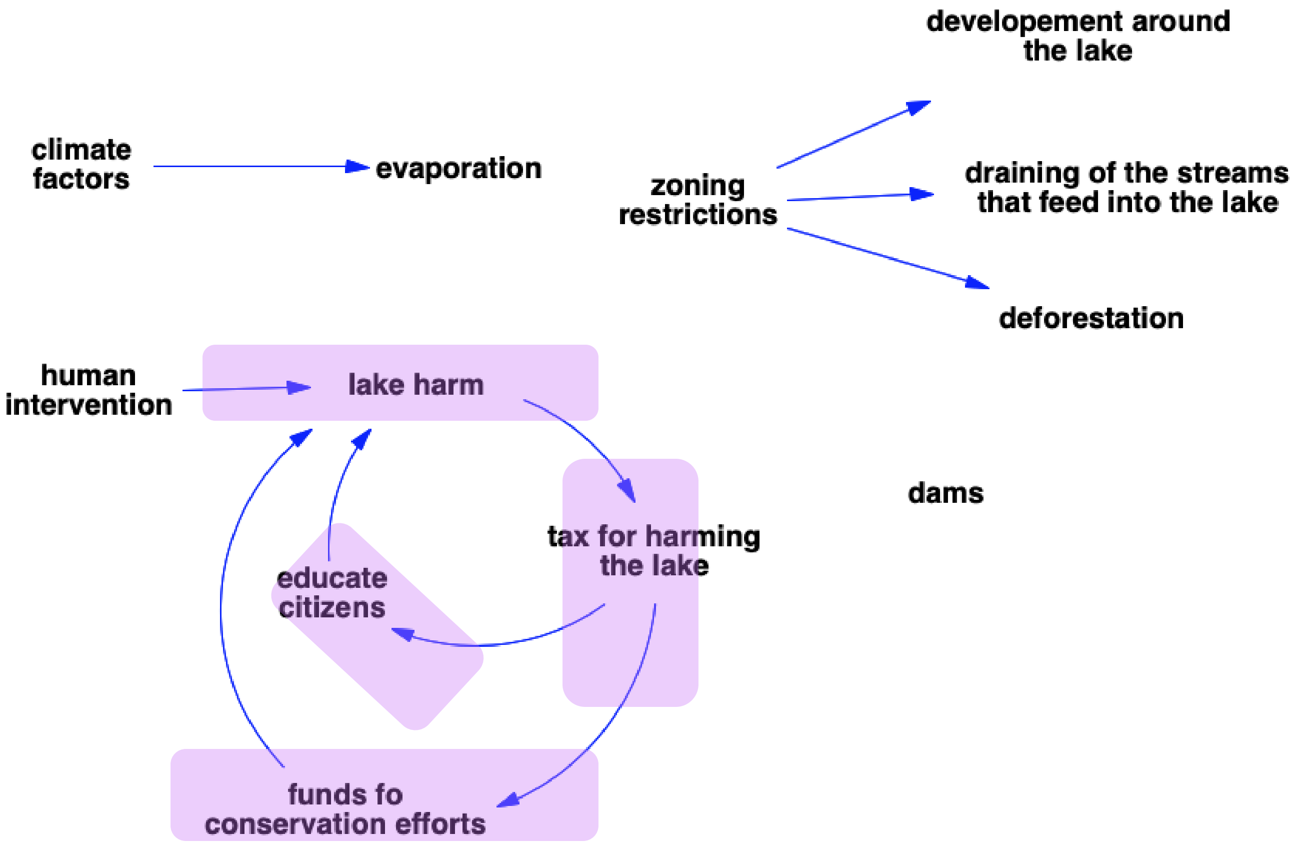
Measure	#	Type of ST
Variables	11	✓ Detailed complexity (Richardson, 1994) ✓ Dynamic thinking (Stave & Hopper, 2007)
Causal links	10	✓ Interconnectivity (Dorani et al., 2015) ✓ Cause-effect thinking (Stave & Hopper, 2007)
Closed loops		✓ System-as-cause thinking (Dorani et al., 2015) ✓ Closed-loop thinking (Richardson, 1994) ✓ Identifying feedback (Stave & Hopper, 2007)
Middle Nodes		✓ Operational thinking (Haque et al., 2023)
Connectivity		✓ Cyclomatic complexity (Naugle et al., 2021) ✓ Link density (Plate, 2010)

3. ST Measure – Number of Closed Loops



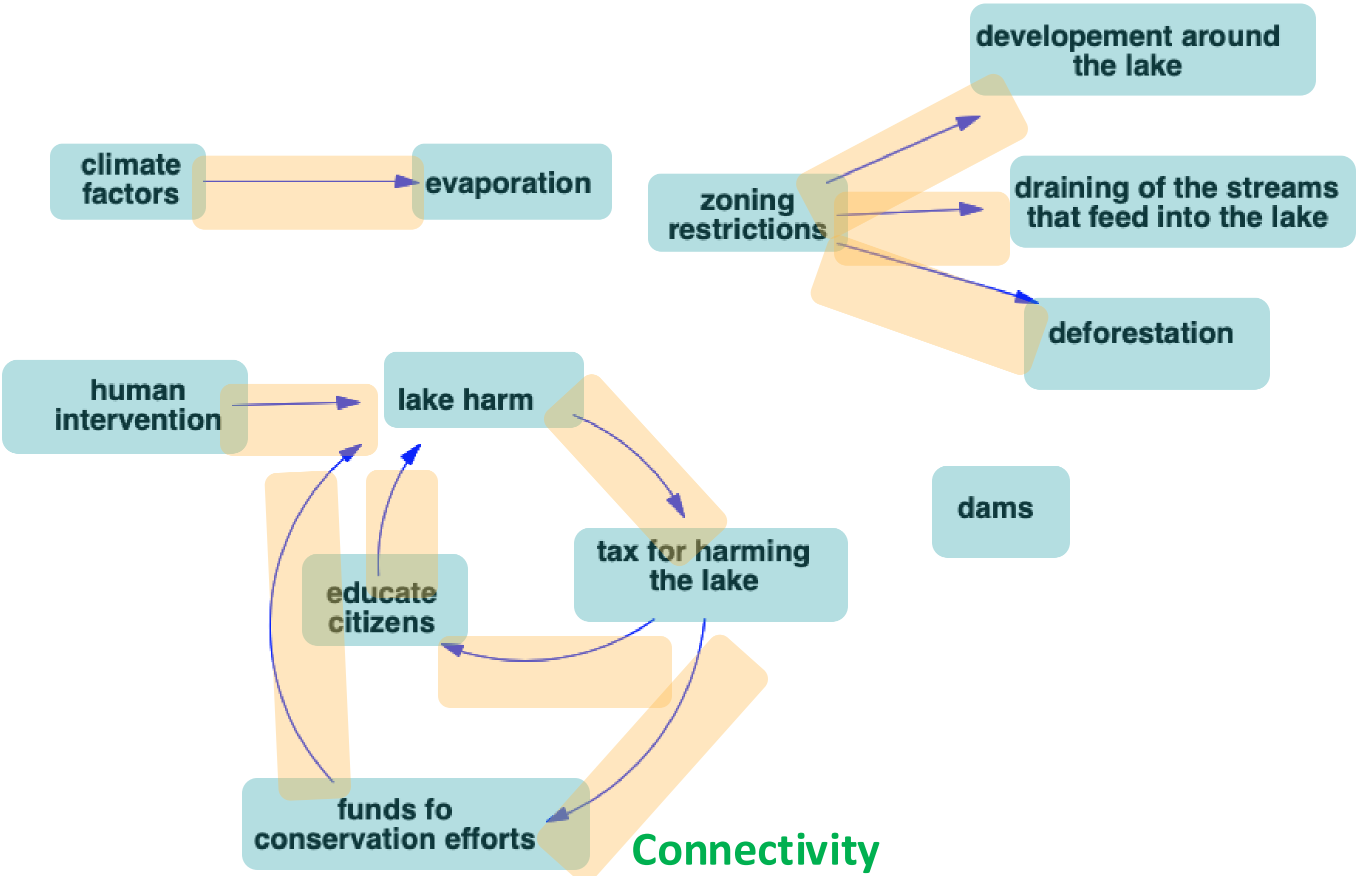
Measure	#	Type of ST
Variables	11	✓ Detailed complexity (Richardson, 1994) ✓ Dynamic thinking (Stave & Hopper, 2007)
Causal links	10	✓ Interconnectivity (Dorani et al., 2015) ✓ Cause-effect thinking (Stave & Hopper, 2007)
Closed loops	2	✓ System-as-cause thinking (Dorani et al., 2015) ✓ Closed-loop thinking (Richardson, 1994) ✓ Identifying feedback (Stave & Hopper, 2007)
Middle Nodes		✓ Operational thinking (Haque et al., 2023)
Connectivity		✓ Cyclomatic complexity (Naugle et al., 2021) ✓ Link density (Plate, 2010)

4. ST Measure – Number of Middle Nodes



Measure	#	Type of ST
Variables	11	✓ Detailed complexity (Richardson, 1994) ✓ Dynamic thinking (Stave & Hopper, 2007)
Causal links	10	✓ Interconnectivity (Dorani et al., 2015) ✓ Cause-effect thinking (Stave & Hopper, 2007)
Closed loops	2	✓ System-as-cause thinking (Dorani et al., 2015) ✓ Closed-loop thinking (Richardson, 1994) ✓ Identifying feedback (Stave & Hopper, 2007)
Middle Nodes	4	✓ Operational thinking (Haque et al., 2023)
Connectivity		✓ Cyclomatic complexity (Naugle et al., 2021) ✓ Link density (Plate, 2010)

5. ST Measure – Connectivity



Connectivity

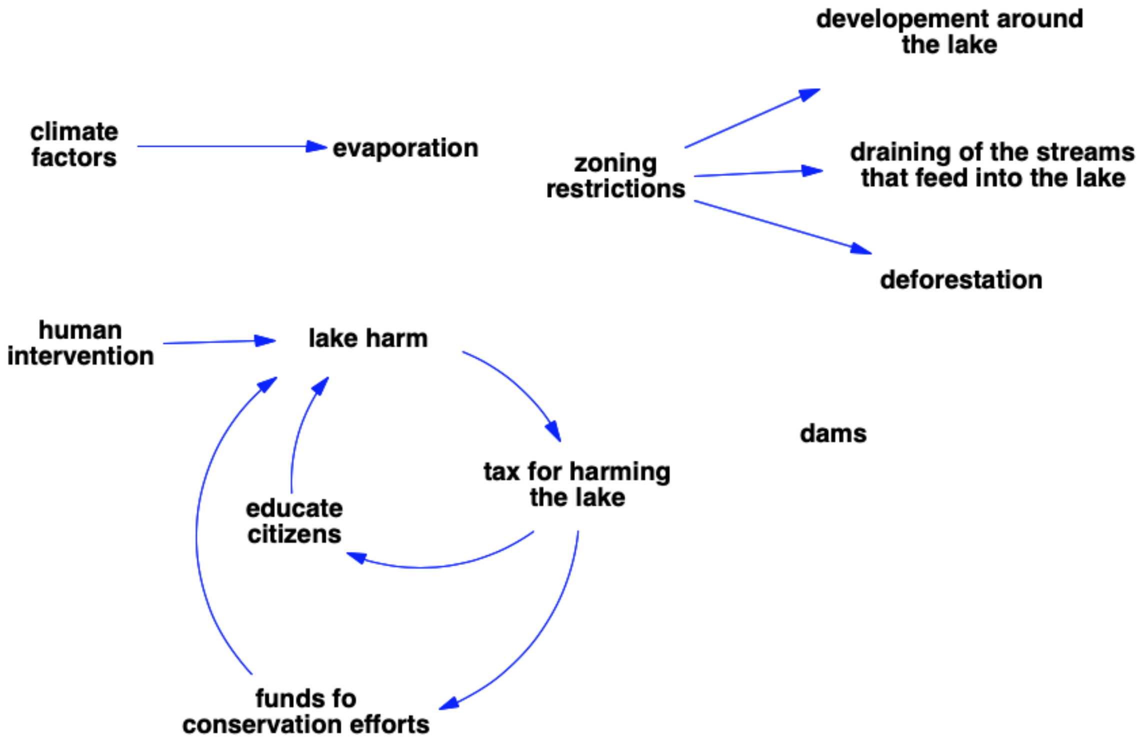
= # of variables – # of casual Links

= 11 – 10 = 1

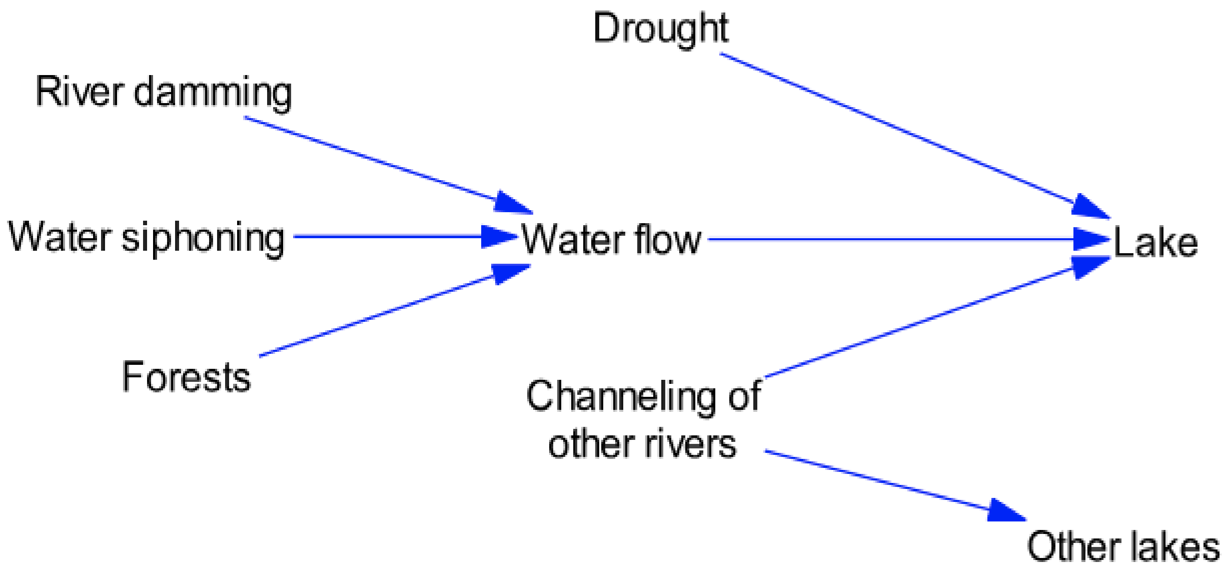
Measure	#	Type of ST
Variables	11	✓ Detailed complexity (Richardson, 1994) ✓ Dynamic thinking (Stave & Hopper, 2007)
Causal links	10	✓ Interconnectivity (Dorani et al., 2015) ✓ Cause-effect thinking (Stave & Hopper, 2007)
Closed loops	2	✓ System-as-cause thinking (Dorani et al., 2015) ✓ Closed-loop thinking (Richardson, 1994) ✓ Identifying feedback (Stave & Hopper, 2007)
Middle Nodes	4	✓ Operational thinking (Haque et al., 2023)
Connectivity	1	✓ Cyclomatic complexity (Naugle et al., 2021) ✓ Link density (Plate, 2010)

Varied Perspectives on the Lake Urmia Vignette (LUV)

Participant Response Example A

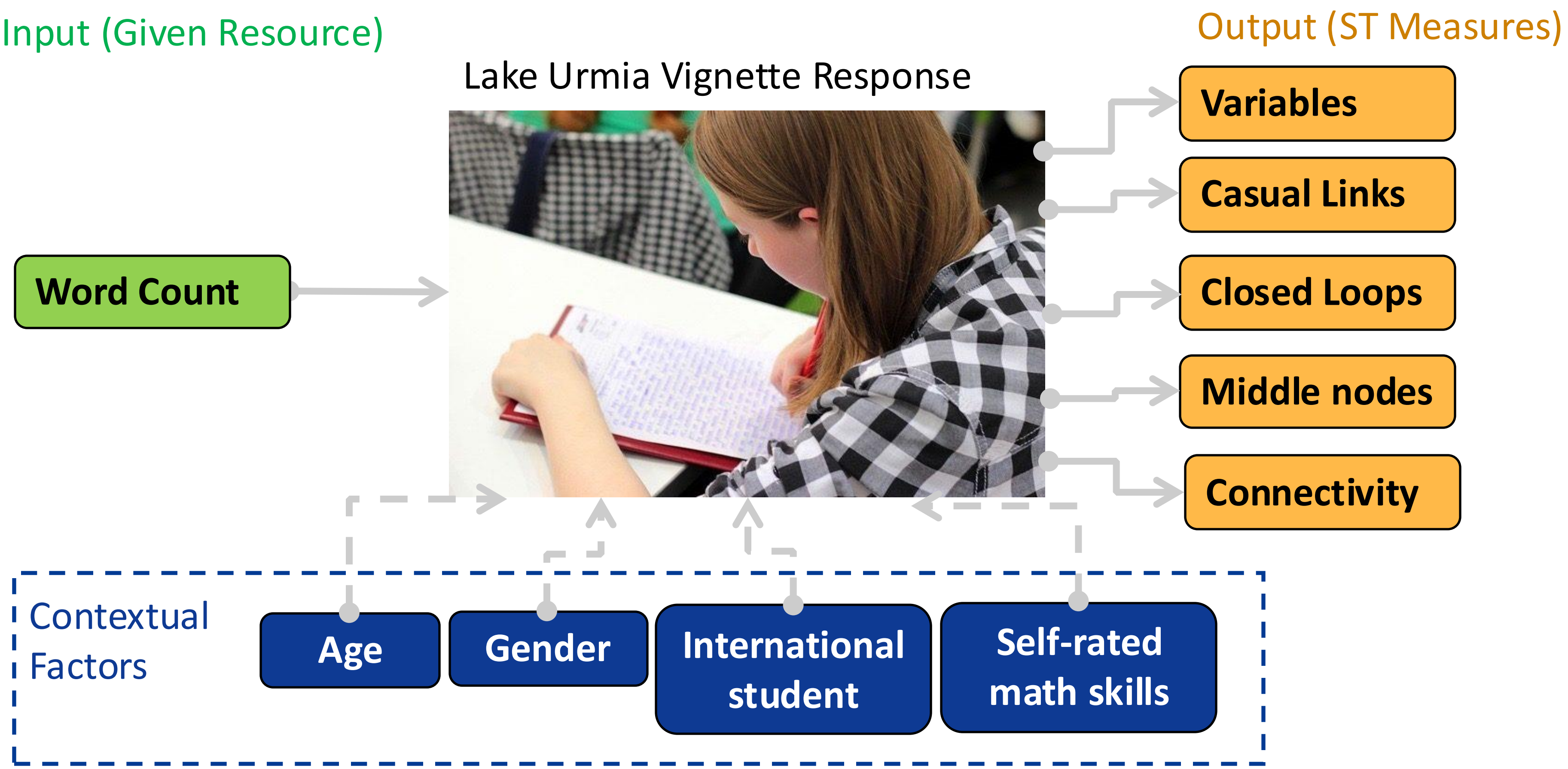


Participant Response Example B



Haque et al. (2023) Mental models, cognitive maps, and the challenge of quantitative analysis of their network representations. System Dynamics Review.

Multidimensional Comprehension Index of Systems Thinking (MIST)



Input (Given Resource)

Lake Urmia Vignette Response

Output (ST Measures)

Word Count

Variables

Casual Links

Closed Loops

Middle nodes

Connectivity

Contextual Factors

Age

Gender

International student

Self-rated math skills

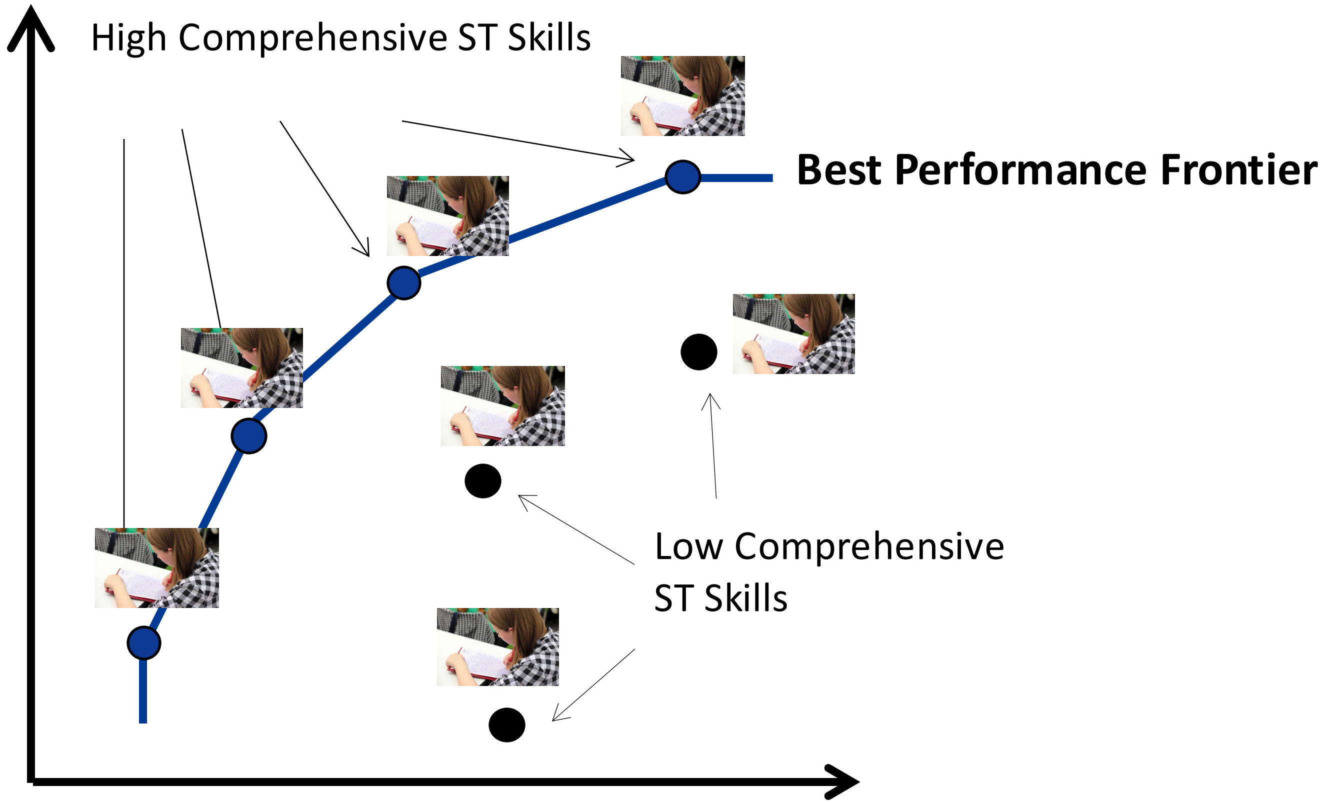
- ✓ Core idea: Benchmarking the Multi-dimensionality of ST through Data Envelopment Analysis (DEA)
- ✓ DEA is a method for **benchmarking the “relative” performance** among LUV responses
- ✓ 144 first-year engineering undergraduate students
- ✓ 30 graduate students

Benchmarking the Comprehensive ST Skills Through DEA

Given the amount of **Input** how much **Output** can one **LUV response** generate?

Output Variables

- # of Variables
- # of Casual Link
- # of Middle Nodes
- # of Connectivity

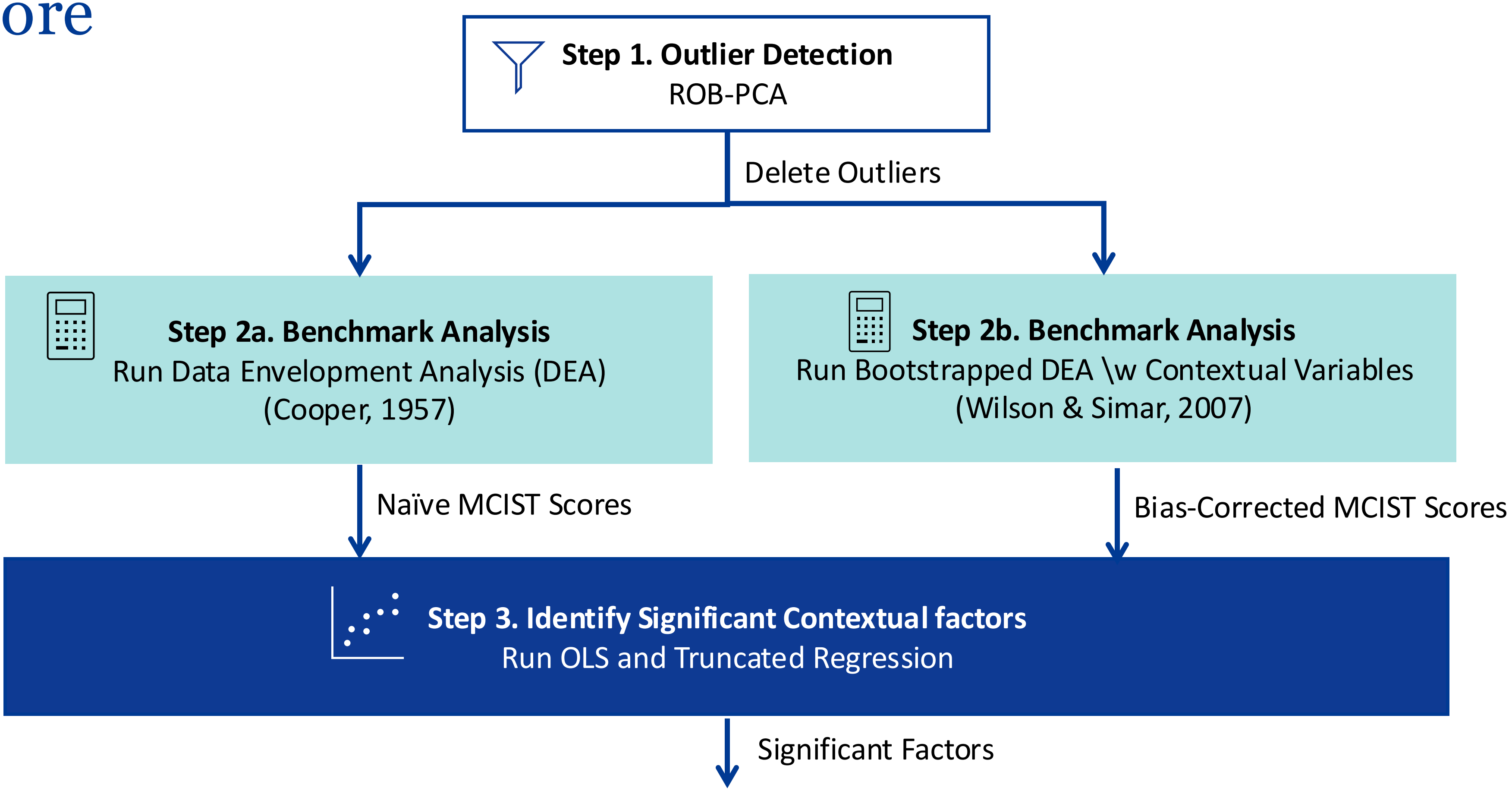

High Comprehensive ST Skills
Best Performance Frontier
Low Comprehensive ST Skills

What do we get?
MIST Score: an index score that represents one’s comprehensive ST skill.

Input Variables

- # of Word Count

Benchmarking Process and the Impact of Contextual Factors on MIST Score



Step 1. Outlier Detection
ROB-PCA

Delete Outliers

Step 2a. Benchmark Analysis
Run Data Envelopment Analysis (DEA)
(Cooper, 1957)

Step 2b. Benchmark Analysis
Run Bootstrapped DEA \w Contextual Variables
(Wilson & Simar, 2007)

Naïve MCIST Scores

Bias-Corrected MCIST Scores

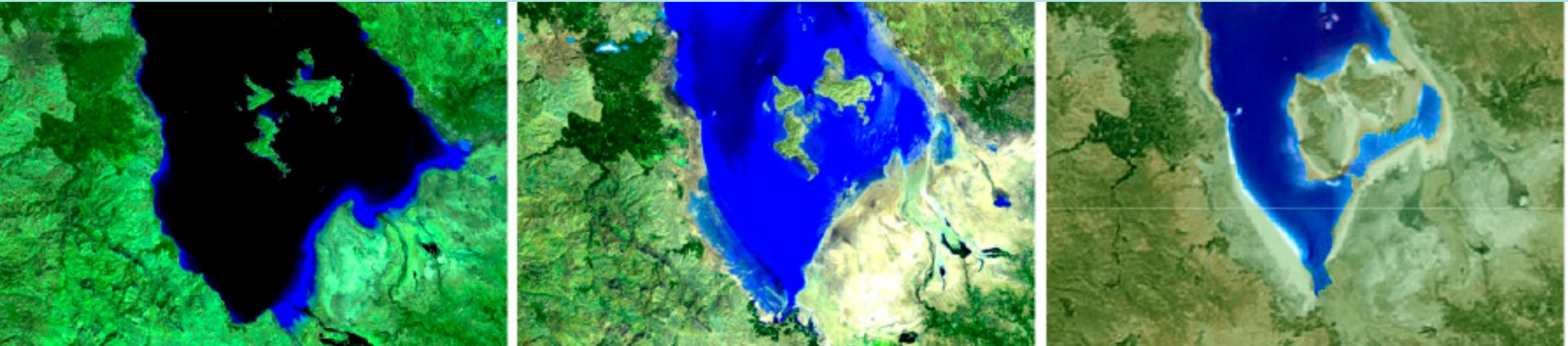
Step 3. Identify Significant Contextual factors
Run OLS and Truncated Regression

Significant Factors

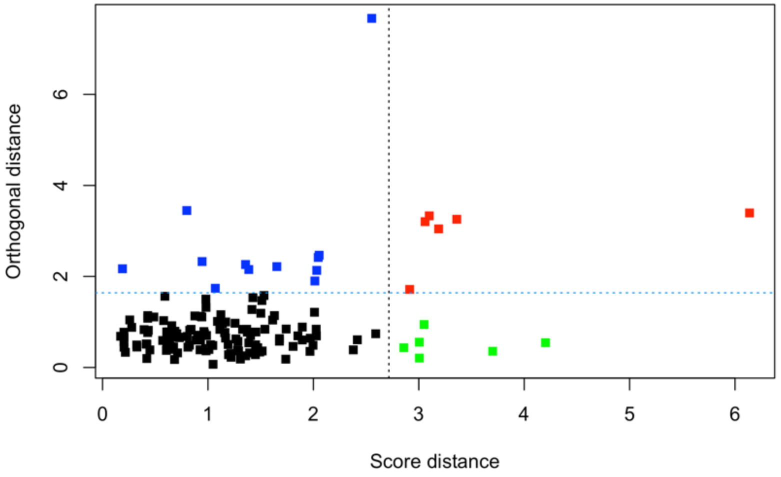
✓ This framework evaluates the **impact of contextual factors** on **MCIST scores**, enabling a comprehensive understanding of individuals' ST performance.

Results

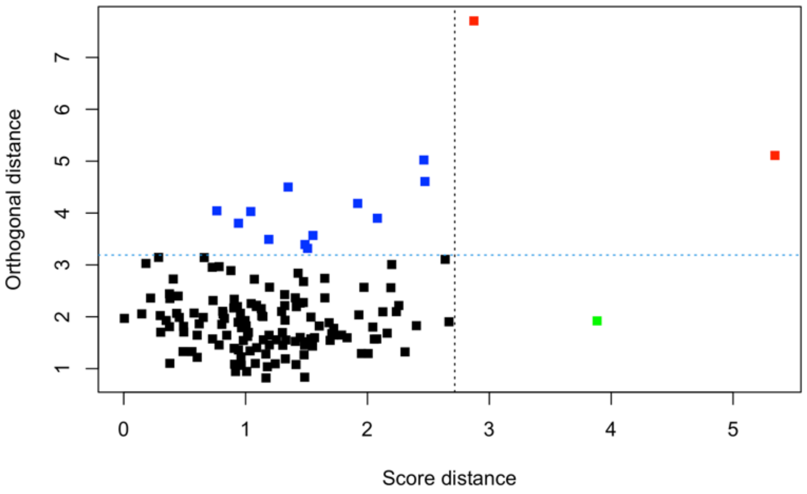
- Comprehensive ST Skills Performance
- The Impact of Contextual Variables on MIST scores



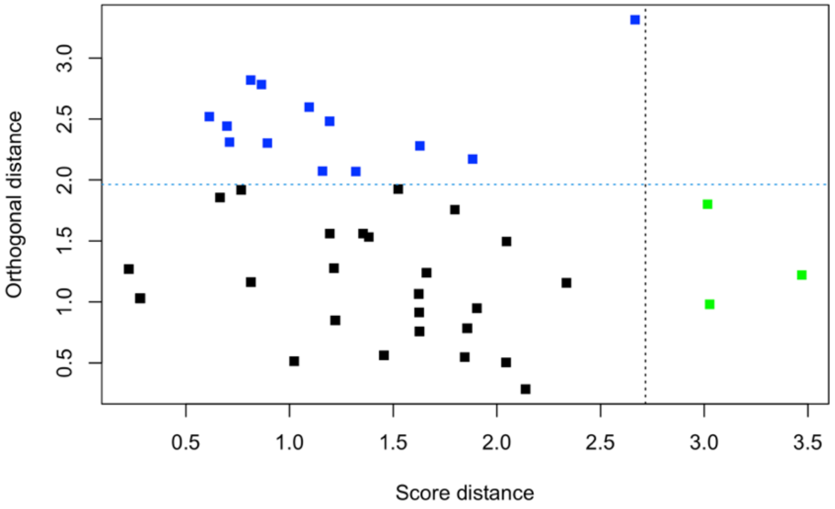
Step 1. Outlier Detection - ROB-PCA



ROBPCA for outlier identifications on **input** and **output** variables



ROBPCA for outlier identifications on **input, output, and contextual** variables



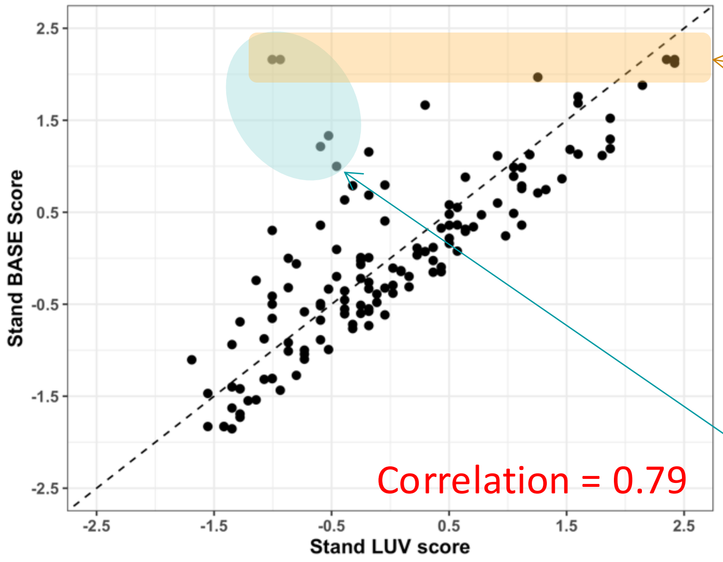
ROBPCA for outlier identifications on **only contextual** variables

- ✓ Drop the **seven outliers** and continued to the next step.

Step 2. Benchmark Analysis - Comprehensive ST Skills Performance

We test on three types of models specification:

Model	Input	Output				
	Word Count	Variables	Links	Loops	Connectivity	Middle Nodes
LUV (Linear)	N/A	✓	✓	✓		
BASE (DEA)	✓	✓	✓	✓		
Multi-index (DEA)	✓		✓	✓	✓	✓

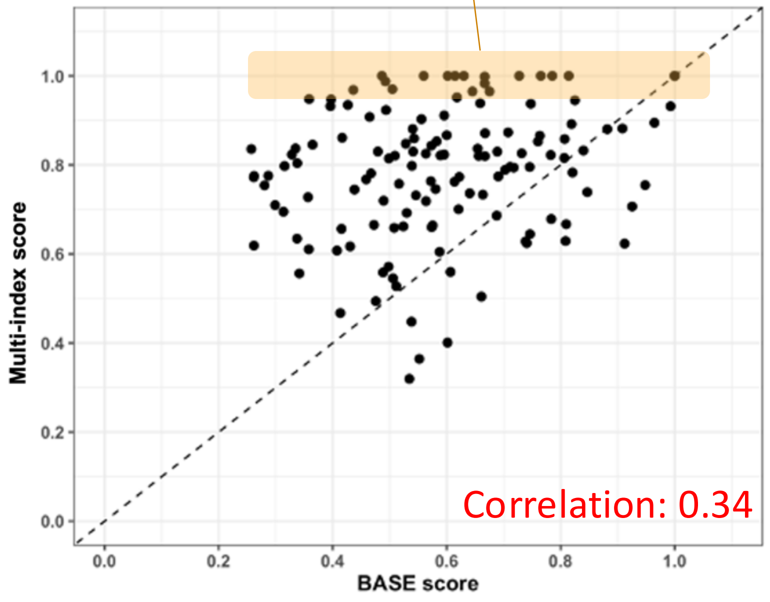


Correlation = 0.79

High ST skills based on the BASE model

The BASE model **does not penalize** responses with a lower Word Count compared to the LUV score.

Increase of # of High ST skill Responses due to adding “Connectivity” and “Middle Nodes”



Correlation: 0.34

Step 3. The Impact of Contextual Variables on MIST

Undergraduate student dataset (N=137)

Contextual Variables	LUV score		Naïve Multi-index score		Bias-corrected Multi-index score	
	Coeff.	Std.	Coeff.	Std.	Coeff.	Std.
(Intercept)	38.961	21.204	0.617	0.469	4.231	5.626
Gender	-0.555	1.380	0.021	0.031	-1.594	1.823
International	8.540	4.213*	0.083	0.093	-0.239	0.312
Age	-1.304	1.123	0.009	0.025	-0.422	0.413
Self-Rate Math	0.833	0.848	0.003	0.019	0.033	0.219
Feedback Score	1.108	0.867	-0.021	0.019	0.359	0.261
N	135				135	
R2	0.073		0.02		N/A	
Adjusted R2	0.037		-0.02		N/A	

Significant Factors

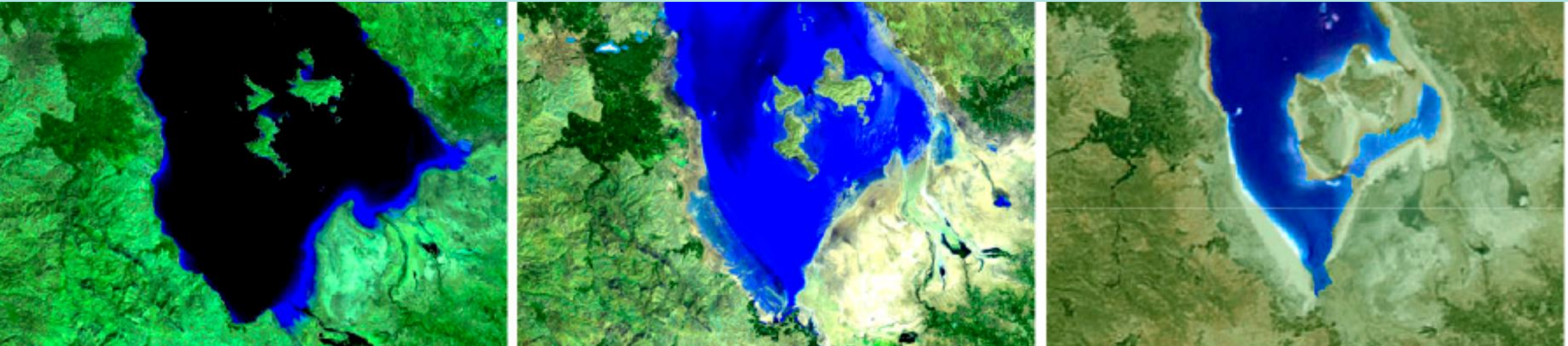
Graduate student dataset (N=27)

Contextual Variables	LUV score		Naïve Multi-index score		Bias-corrected Multi-index score	
	Coeff.	Std.	Coeff.	Std.	Coeff.	Std.
(Intercept)	18.593*	8.47	0.810**	0.212	1.644*	0.795
Level of study	4.791.	2.333	0.021	0.058	0.073	0.19
Gender	0.325	0.211	0.026	0.06	-0.218	0.213
International	6.825*	2.72	0.157*	0.068	-0.517*	0.255
Age	0.325	0.211	0.011.	0.005	-0.049.	0.03
Training In Systems	-3.177.	1.572	-0.03	0.039	0.119	0.131
Self-Rate Math	-4.433*	1.86	-0.093.	0.047	0.281.	0.165
Feedback Score	1.809	1.245	0.036	0.031	-0.116	0.093
N	27		27		27	
R2	0.439		0.341		N/A	
Adjusted R2	0.232		0.099		N/A	

Note. “*” indicates p < .05. Binary variables Female = 1, International = 1.

Conclusions

Lesson Learnt, limitations and future research



Conclusion

Overcoming the Complexity of Text-Based ST Assessment:

- ✓ Addressed the challenge of measuring students' ST skills from text-based responses.
- ✓ Developed a multidimensional index to evaluate ST skills by assessing various ST components, including Variables, Casual links, Loops, Middle Nodes, and Connectivity.

The Impact of Contextual Factors on MIST Score:

- ✓ No significant predictors of one's level of ST in the undergraduate dataset.
- ✓ International students, self-reported math skills, and older students demonstrate higher levels of ST skills in the graduate student dataset.

Practical Implications:

- ✓ Test on two different datasets (graduate and undergraduate students).
- ✓ Initial validation with LUV: expanding possibilities for broader use.

Limitations and Future Research

- Further research should consider the dynamic changes in individuals' understanding of complexity.

- Test on other DEA models with different fundamental assumptions, such as the free disposal hull model, additive, and slack-based models (Deprins & Simar, 1984; Ali & Seiford, 1993; Tone, 2001).

- Assessment of individuals' ST skills as defined by other schools of thought is a future avenue of research. Invite researchers to examine other schools of thought, including critical ST, soft ST, and general system theory (Von Bertalanffy, 1973; Checkland & Haynes, 1994 & Jackson, 2016).

- Improvement of the coding procedure of translating textual data to mental Maps.

Paper II

Leveraging Large Language Models for Automated Causal Loop Diagram Generation

= (mod - r) % mod; }
r) % mod; } } } while
(0); done ; if (fj < N
A[i][j] = (A[i][j] - A
depth (N), rt (N), pos (N), tree
; dfsSz (u); siz [v] += siz [u];
[0] ? rt [v] : u); dfsHld (u);

Outline

Introduction

- Translation from Textual Data to Mental Maps
- Prompt Engineering in Large Language Models
- Research Question

Data & Methods

- Experiment Setup
- Prompt Engineering for Casual Loop Diagram Generation

Results

- Case 1 – Single Reinforcing Loop
- Case 2 – Two Balancing Loop
- Case 3 – Two Balancing Loops with Exogenous Variables

CLD Generator Demo

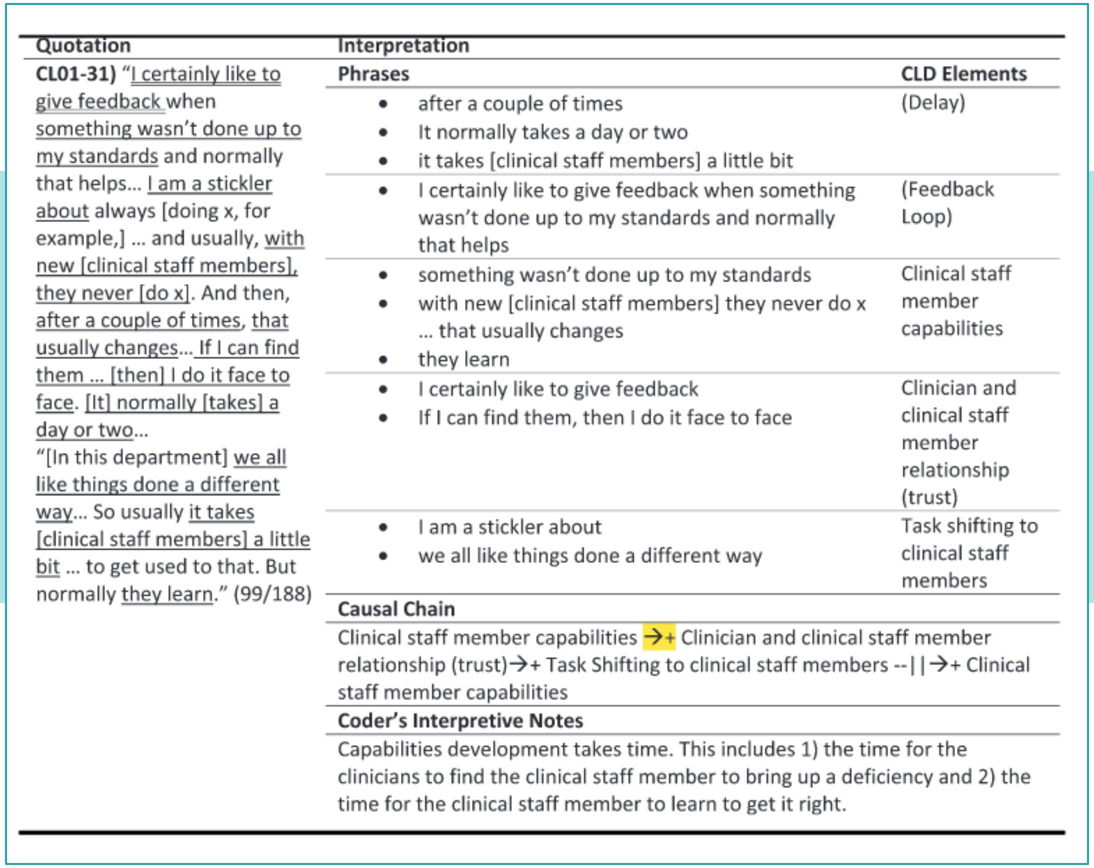
Conclusions

- Lesson Learnt
- Limitations and Future Research

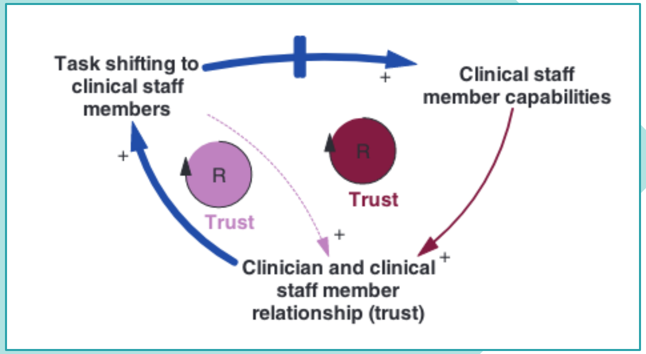
Introduction

- ➢ Constructing Mental Maps from Textual data
- ➢ Prompt Engineering in Large Language Models
- ➢ Research Questions

Constructing Mental Maps from Textual data has been a Manually or Semi-automated Process

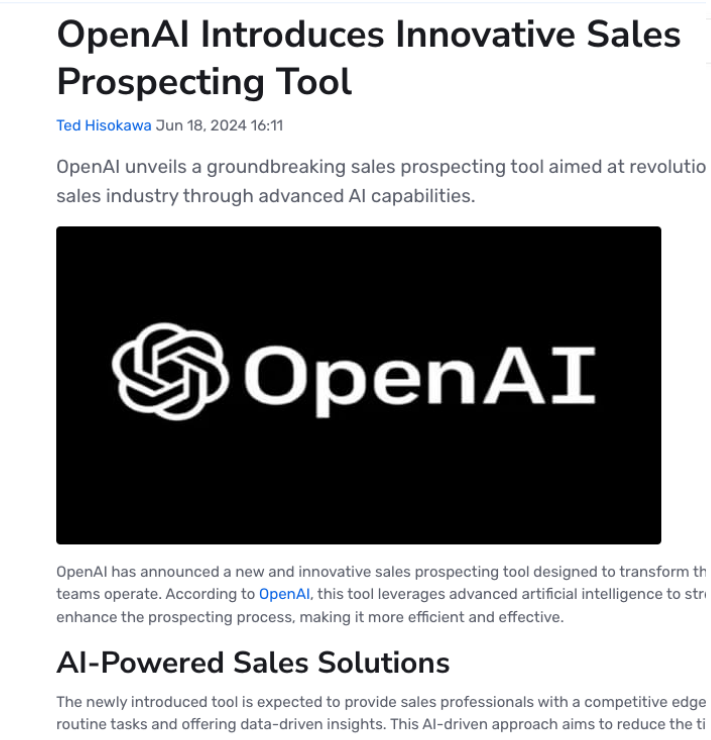


Quotation	Interpretation Phrases	CLD Elements
CL01-31) "I certainly like to give feedback when something wasn't done up to my standards and normally that helps… I am a stickler about always [doing x, for example,] … and usually, with new [clinical staff members], they never [do x]. And then, after a couple of times, that usually changes… If I can find them … [then] I do it face to face. [It] normally [takes] a day or two… "[In this department] we all like things done a different way… So usually it takes [clinical staff members] a little bit … to get used to that. But normally they learn." (99/188)	• after a couple of times • It normally takes a day or two • it takes [clinical staff members] a little bit	(Delay)
	• I certainly like to give feedback when something wasn't done up to my standards and normally that helps	(Feedback Loop)
	• something wasn't done up to my standards • with new [clinical staff members] they never do x … that usually changes • they learn	Clinical staff member capabilities
	• I certainly like to give feedback • If I can find them, then I do it face to face	Clinician and clinical staff member relationship (trust)
	• I am a stickler about • we all like things done a different way	Task shifting to clinical staff members
	Causal Chain Clinical staff member capabilities →+ Clinician and clinical staff member relationship (trust)→+ Task Shifting to clinical staff members --||→+ Clinical staff member capabilities	
	Coder's Interpretive Notes Capabilities development takes time. This includes 1) the time for the clinicians to find the clinical staff member to bring up a deficiency and 2) the time for the clinical staff member to learn to get it right.	


Task shifting to clinical staff members
+
Clinical staff member capabilities
+
R
Trust
R
Trust
+
+
Clinician and clinical staff member relationship (trust)
+

Tomoaia-Cotisel et al. (2022) Rigorously interpreted quotation analysis for evaluating causal loop diagrams in Late-Stage conceptualization. System Dynamics Review, 38(1), 41-80.

With the advancement of Generative AI, many manual tasks can be enhanced using LLMs



OpenAI Introduces Innovative Sales Prospecting Tool

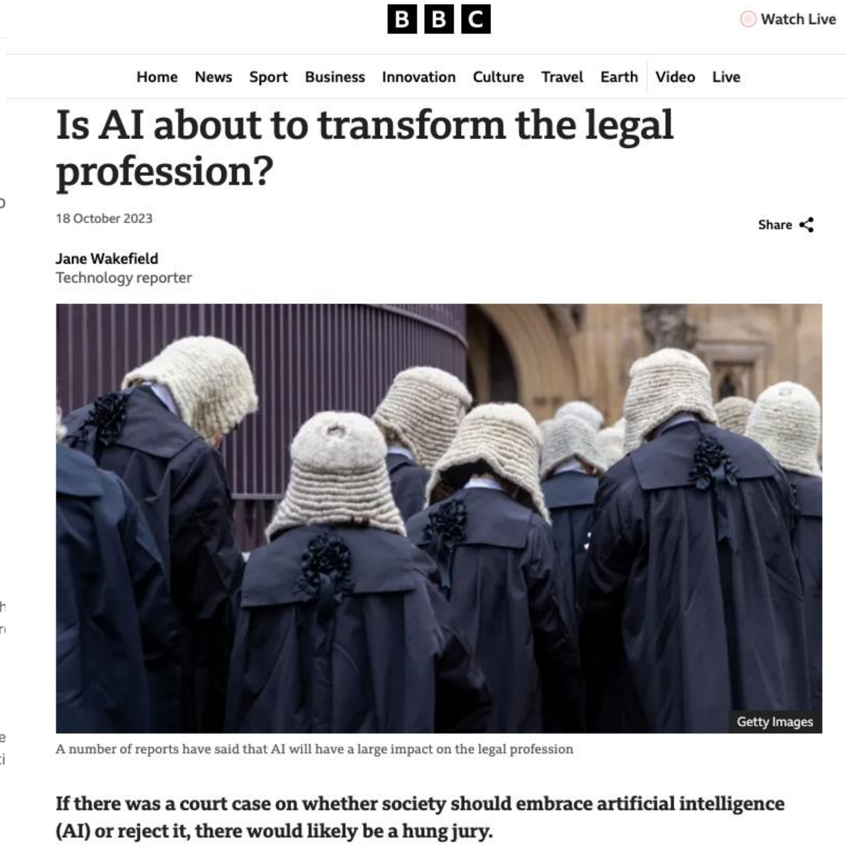
Ted Hisokawa Jun 18, 2024 16:11

OpenAI unveils a groundbreaking sales prospecting tool aimed at revolutio sales industry through advanced AI capabilities.

OpenAI has announced a new and innovative sales prospecting tool designed to transform th teams operate. According to OpenAI, this tool leverages advanced artificial intelligence to str enhance the prospecting process, making it more efficient and effective.

AI-Powered Sales Solutions

The newly introduced tool is expected to provide sales professionals with a competitive edge routine tasks and offering data-driven insights. This AI-driven approach aims to reduce the ti



Is AI about to transform the legal profession?

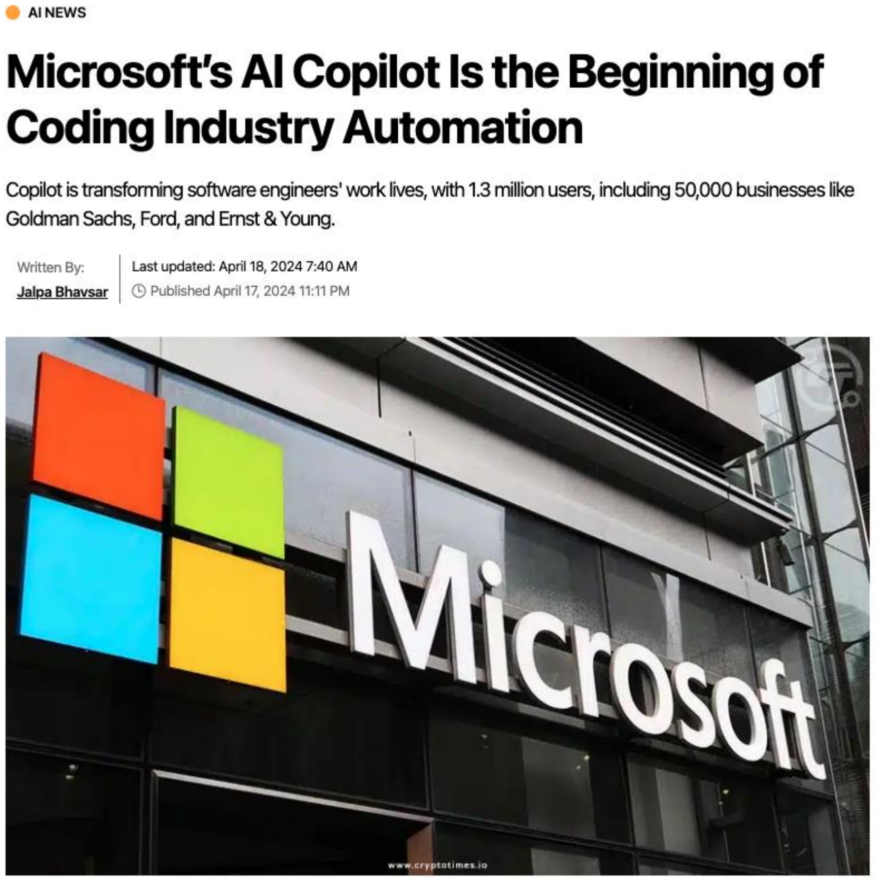
18 October 2023

Jane Wakefield
Technology reporter

Getty Images

A number of reports have said that AI will have a large impact on the legal profession

If there was a court case on whether society should embrace artificial intelligence (AI) or reject it, there would likely be a hung jury.



AI NEWS

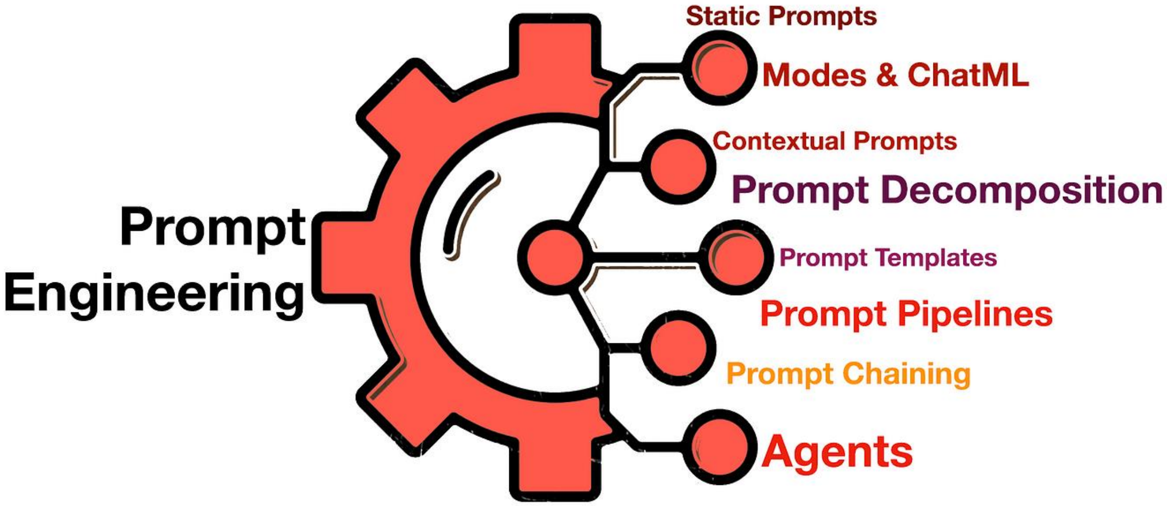
Microsoft's AI Copilot Is the Beginning of Coding Industry Automation

Copilot is transforming software engineers' work lives, with 1.3 million users, including 50,000 businesses like Goldman Sachs, Ford, and Ernst & Young.

Written By: Jalpa Bhavsar

Last updated: April 18, 2024 7:40 AM
Published April 17, 2024 11:11 PM

Prompt Engineering in Large Language Models



Advantages of Prompt Engineering:

- ✓ Improving Precision in Predictions
- ✓ Efficient Information Extraction
- ✓ Controlling Model Behavior
- ✓ Reducing Hallucination

Brown et al. (2020) Language Models are Few-Shot Learners.
Liu et al. (2021) Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing.

Research Questions

1. Can we automatically translate text into Casual Loop Diagrams (CLDs)?

2. Can we improve the quality of the generated CLDs with prompt engineering?

Data & Methods

- Experiment Setup
- Prompt Engineering for CLD Generation

Experiment Setup

LLM Model Selection:

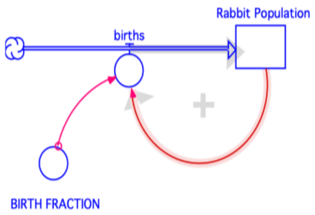
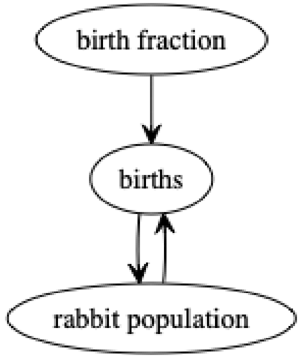
- OpenAI's text-davinci-003 model

Dataset:

- 44 CLDs containing 1-4 feedback loops.
- Text-CLD pairs from leading SD publications.

Example Text-CLD pair provided to the LLM

Input: Dynamic Hypothesis	Output: Causal Loop Diagram	Output: CLD – Digraph String format	Output: CLD - Digraph

The larger the population, the greater the number of births. increases, the faster the population increases. The more the birth rate increases, the faster the population increases.		digraph {
		"births" -> "rabbit population" [arrowhead = vee]
		"rabbit population"->"births"[arrowhead = vee]
		"birth fraction" -> "births"[arrowhead = vee]
		}
			

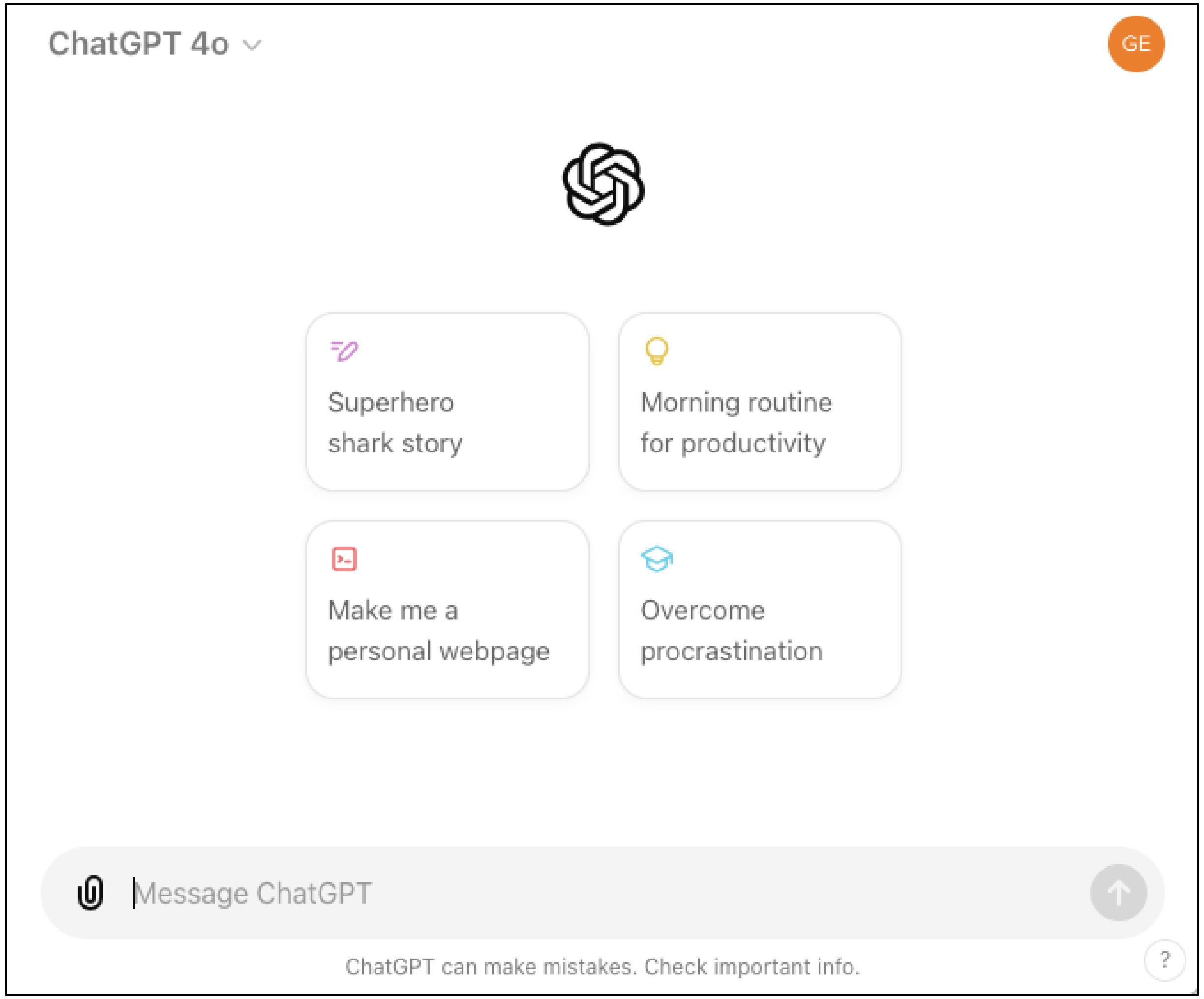
Note: A positive polarity between two variables is represented as arrowhead = vee **(->),** while a negative polarity is represented as arrowhead = tee **(-|).**

Testing Types of Prompting Techniques for CLD Generation

Descriptions of 4 types of prompting techniques

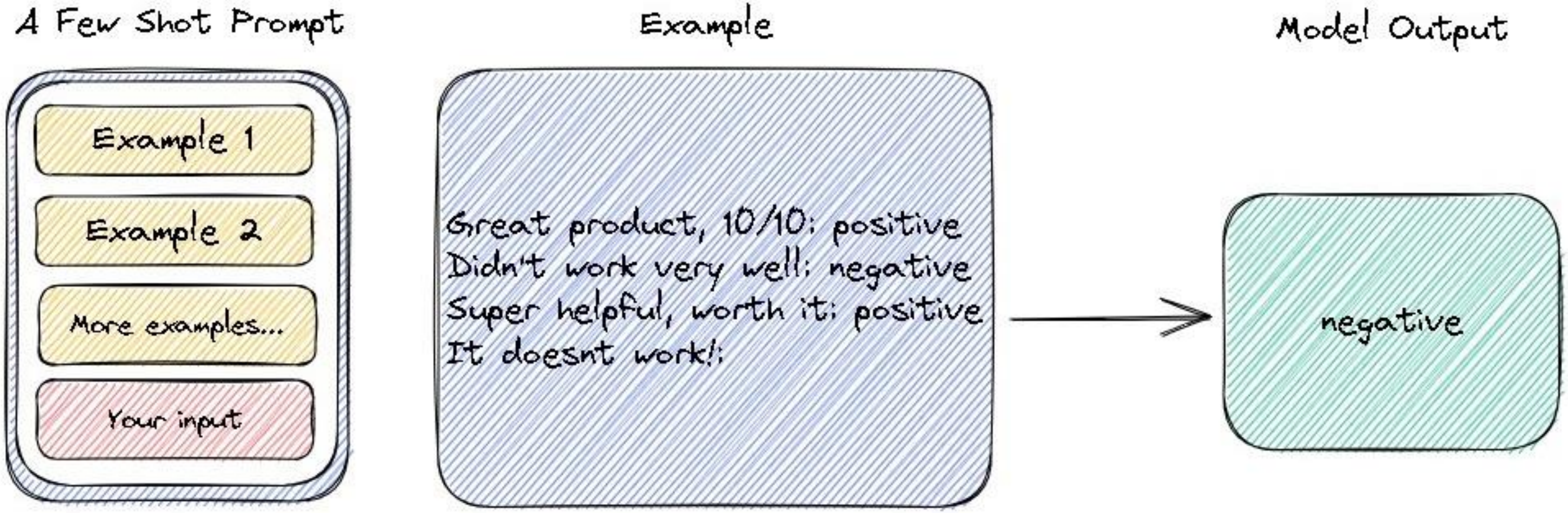
Approach		Description of the Method
1	Zero-shots learning	Baseline. No prior examples are given.
2	Few-shots learning	Given a few examples to LLMs in advanced.
3	Guided Prompts	Incorporated curated prompts, i.e., specific instructions to guide the model’s response.
4	Two-stage few-shot learning	Mimics the thought process of a human SD modeler, focusing first on variable identification and subsequently on mapping out causal relationships.

Approach 1: Zero-shot Learning



ChatGPT 4o
GE
Superhero shark story
Morning routine for productivity
Make me a personal webpage
Overcome procrastination
Message ChatGPT
ChatGPT can make mistakes. Check important info.
?

Approach 2: Few Shots Learning



Source: https://www.linkedin.com/pulse/few-shot-prompting-aris-ihwan

Approach 3: Guided Prompts

Guided Prompt instructs as follows:

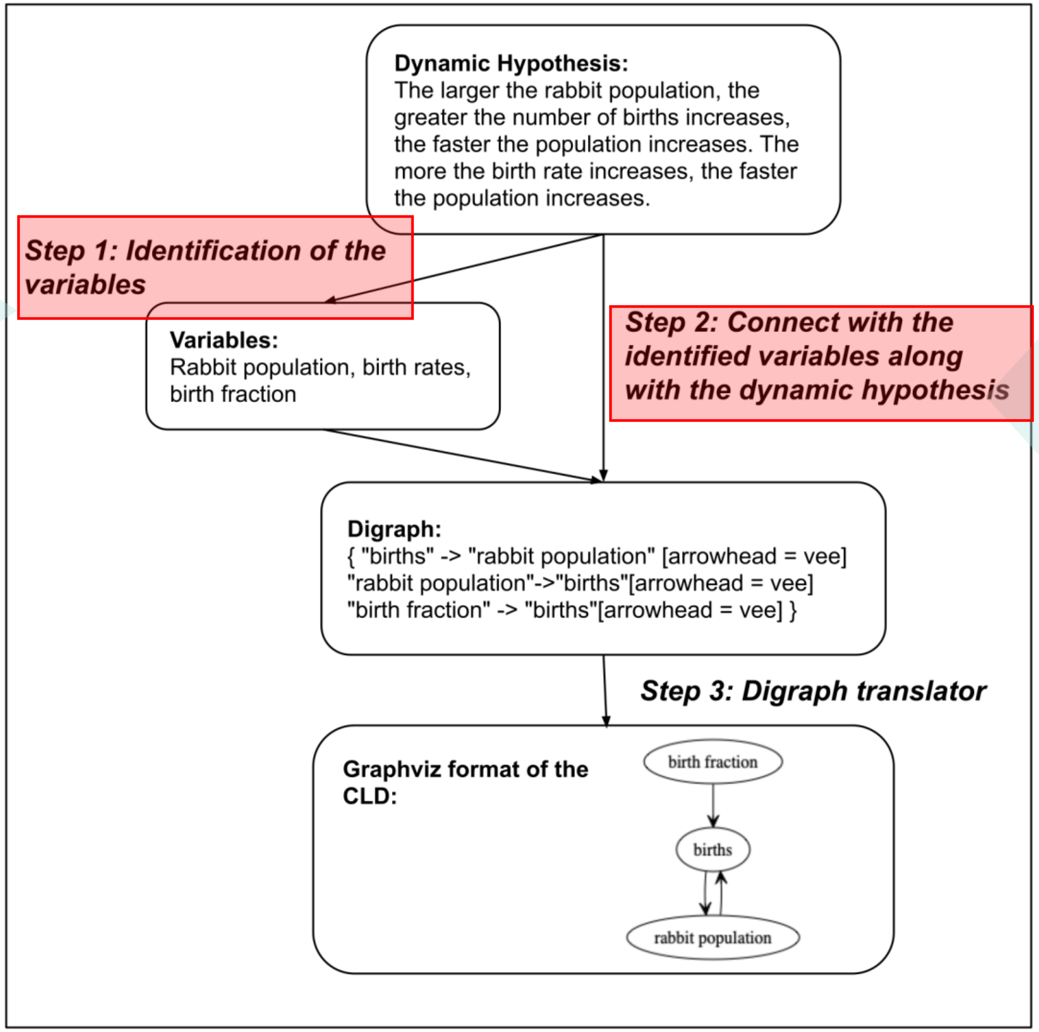
First, Render a list of variable names from the text given. The **variable names should be nouns or nouns phrases**. The variable names should have a sense of directionality. Chose names for which the the meaning of an increase or decrease is clear.

Second, Render **a DOT format** based on the variable names. A positive relationship is indicated by an arrow from the first variable to the second variable with the sign [vee]. A negative relationship is indicated by an arrow from the first variable to the second variable with the sign [tee].

Approach 4:Two-stage few-shot learning

Render a list of variable names from the text given. Following the rules below =

1. *The variable names should be nouns or nouns phrases.*
2. *The variable names should have a sense of directionality.*



Dynamic Hypothesis:
The larger the rabbit population, the greater the number of births increases, the faster the population increases. The more the birth rate increases, the faster the population increases.

Step 1: Identification of the variables

Variables:
Rabbit population, birth rates, birth fraction

Step 2: Connect with the identified variables along with the dynamic hypothesis

Digraph:
{ "births" -> "rabbit population" [arrowhead = vee]
"rabbit population"->"births"[arrowhead = vee]
"birth fraction" -> "births"[arrowhead = vee] }

Step 3: Digraph translator

Graphviz format of the CLD:

Flow Diagram of the Model Setup for Two-stage Approach

The variables' names will be rendered in DOT format. The steps are as follows:

Step 1: Identify the cause-effect relationship between variable names given the dynamic hypothesis.

Step 2: [arrowhead=vee] indicates a positive relationship. A negative relationship is indicated by [arrowhead=tee].

Step 3: Create a DOT format based on the cause-effect relationship.

- This method is follows to the system dynamics modeling approach outlined by (Sterman, 2000, p152).

Results

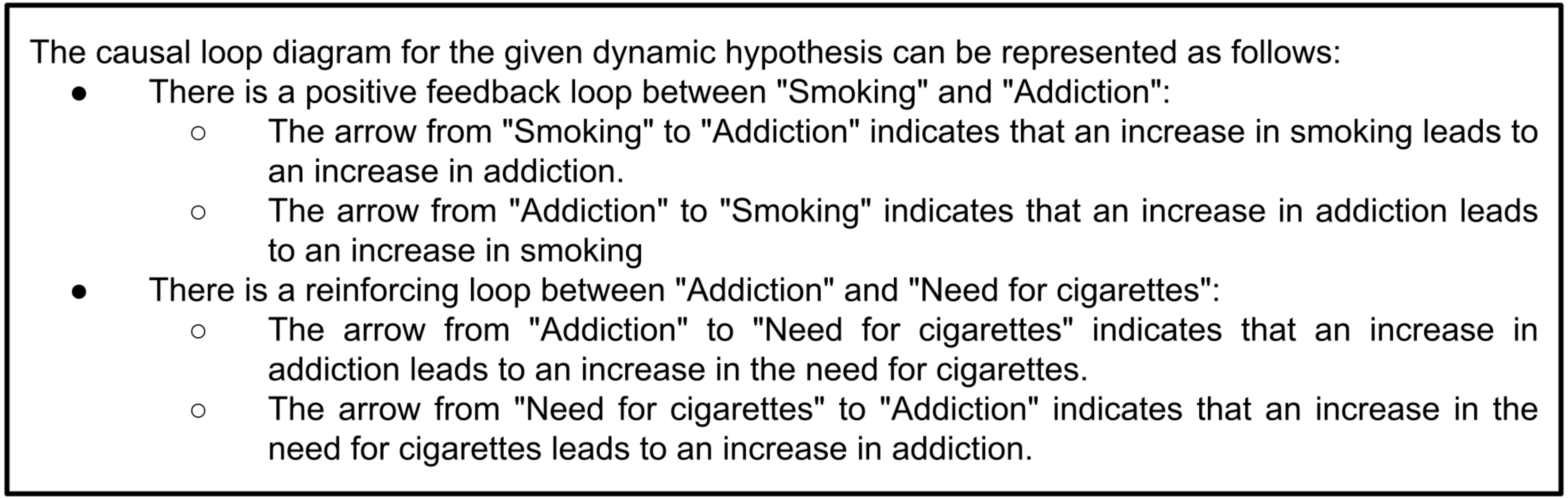
- Case 1 – Single reinforcing loop
- Case 2 – Two Balancing Loops
- Case 3 – Two Balancing Loops with Exogenous Variables

= (mod - r) % mod ; |
r) % mod ; } } } while
(0); done : if (fj < N
A[i][j] = (A[i][j] - A
depth (N), rt (N), pos (N), tree
dfsSz (u); siz [v] += siz [u];
2 rt [v] : u); dfsHld (u);

Case 1: Single reinforcing loop - Smoking cigarettes

INPUT: “The **more** my uncle **smokes**, the **more addicted** he becomes to the nicotine in his cigarettes. After smoking a few cigarettes a long time ago, my uncle began to **develop a need for cigarettes.** The need caused him to smoke even more, which produced an even **stronger need to smoke**. The reinforcing behavior in the addiction process is characteristic of positive feedback.”

LLM Generated Outputs from Zero Shot Approach:

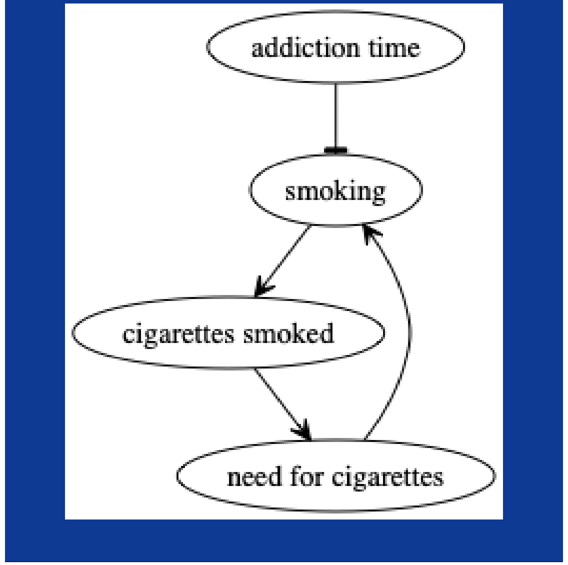


The causal loop diagram for the given dynamic hypothesis can be represented as follows:

- There is a positive feedback loop between "Smoking" and "Addiction":
  - The arrow from "Smoking" to "Addiction" indicates that an increase in smoking leads to an increase in addiction.
  - The arrow from "Addiction" to "Smoking" indicates that an increase in addiction leads to an increase in smoking
- There is a reinforcing loop between "Addiction" and "Need for cigarettes":
  - The arrow from "Addiction" to "Need for cigarettes" indicates that an increase in addiction leads to an increase in the need for cigarettes.
  - The arrow from "Need for cigarettes" to "Addiction" indicates that an increase in the need for cigarettes leads to an increase in addiction.

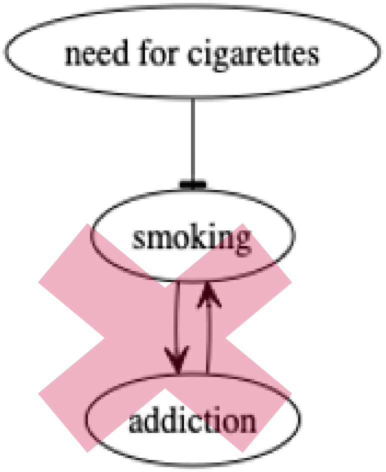
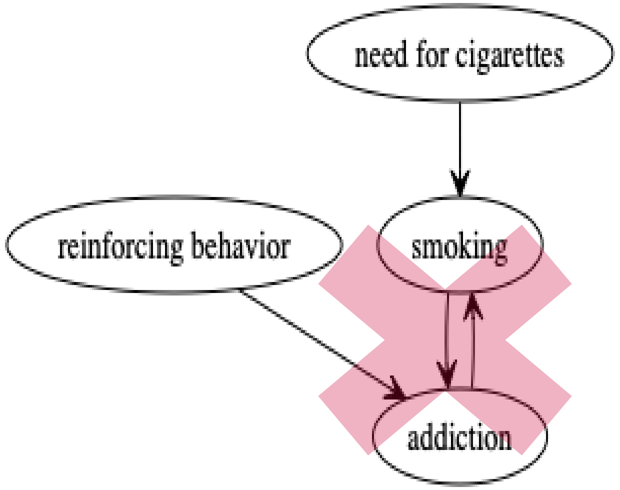
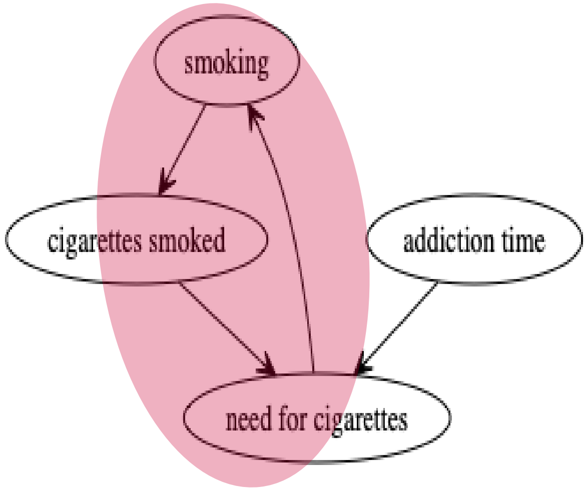
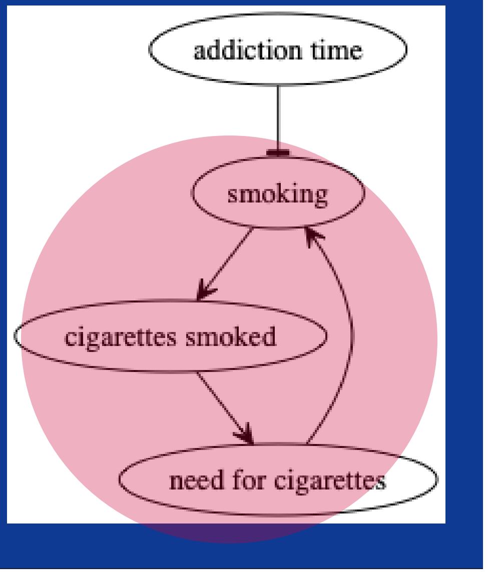
Ground truth (From SD textbook)

Expert CLD



Case 1: Single reinforcing loop - Smoking cigarettes

INPUT: “The **more** my uncle **smokes**, the **more addicted** he becomes to the nicotine in his cigarettes. After smoking a few cigarettes a long time ago, my uncle began to **develop a need for cigarettes.** The need caused him to smoke even more, which produced an even **stronger need to smoke**. The reinforcing behavior in the addiction process is characteristic of positive feedback.”

LLM Generated Outputs			Ground truth (From SD textbook)
Few-shots Approach	**Guided Prompts Approach**	**Two-stage Approach**	**Expert CLD**
			

Note: “->” indicates a positive relationship, “-|” indicates a negative relationship

Case 2: Two Balancing Loops – New Car Inventory

INPUT: "Car production builds the inventory of cars at the dealer. A **higher inventory** leads to a **lower market price**, and **lower market prices** cause **less car production** in the future. If the **price** were to **increase**, the **retail sale of cars** would tend to **fall**. **Retails sales** drain the **inventory of cars** held in stock at the dealership. And **a decline in the inventory** will cause the dealers to **raise** their **prices** in the future."

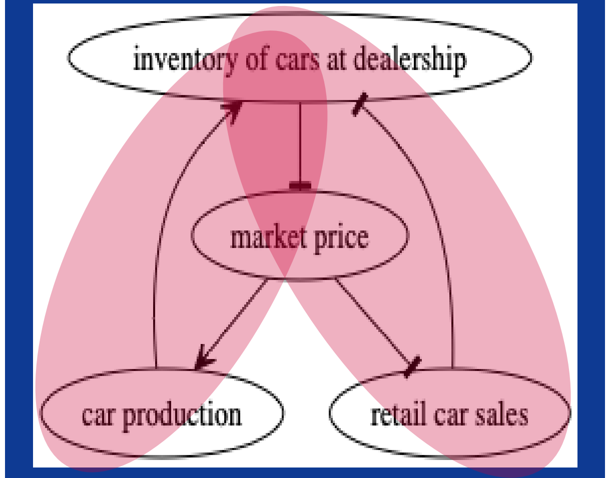
LLM Generated Outputs from Zero Shot Approach:

The causal loop diagram for the given dynamic hypothesis can be represented as follows:

- Car Production --> Inventory of Cars at Dealer --> Market Price --> Car Production
  - ○ Car Production is positively influenced by the Inventory of Cars at Dealer, as a higher inventory allows for more cars to be produced. The Inventory of Cars at Dealer is negatively influenced by the Market Price, as a lower market price leads to a decrease in the inventory.
  - ○ The Market Price is negatively influenced by the Inventory of Cars at Dealer, as a higher inventory leads to a lower market price. Car Production is negatively influenced by the Market Price, as a lower market price leads to a decrease in future car production.
- This causal loop diagram represents a negative feedback loop, where changes in the inventory and market price influence car production, and changes in car production influence the inventory and market price.

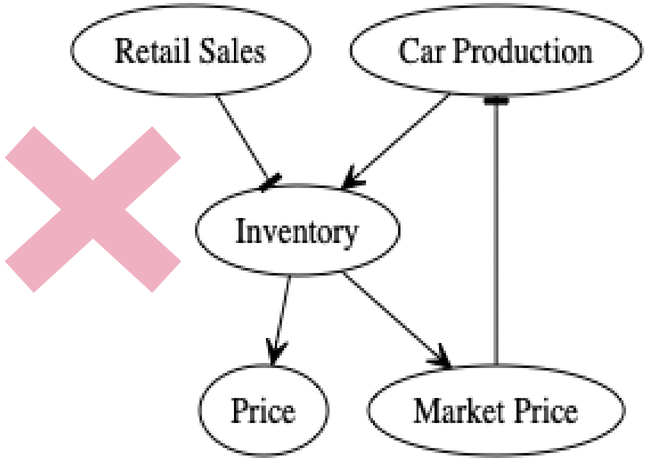
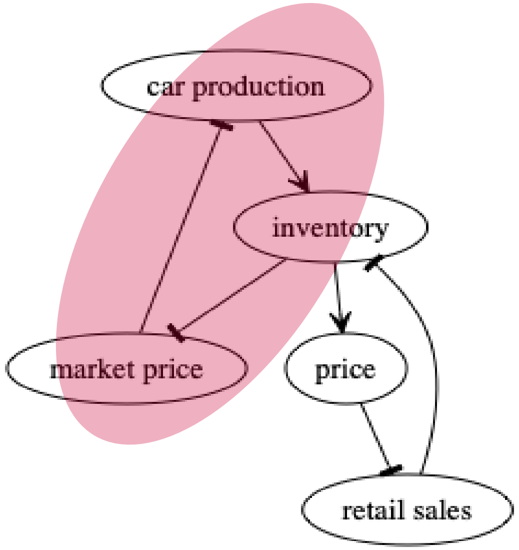
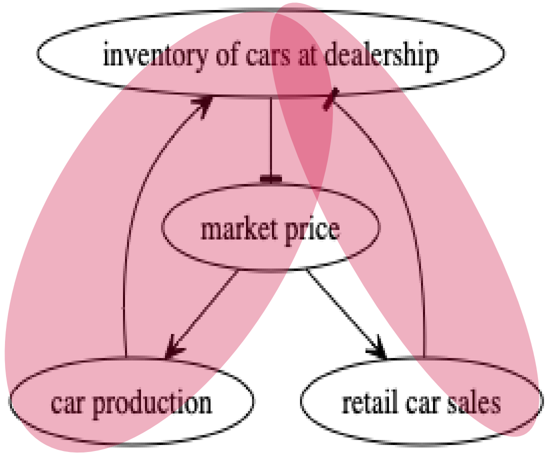
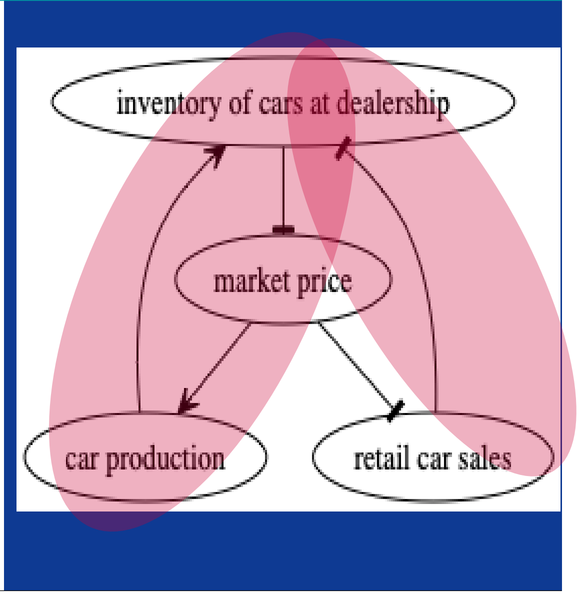
Ground truth (From SD textbook)

Expert CLD



Case 2: Two Balancing Loops – New Car Inventory

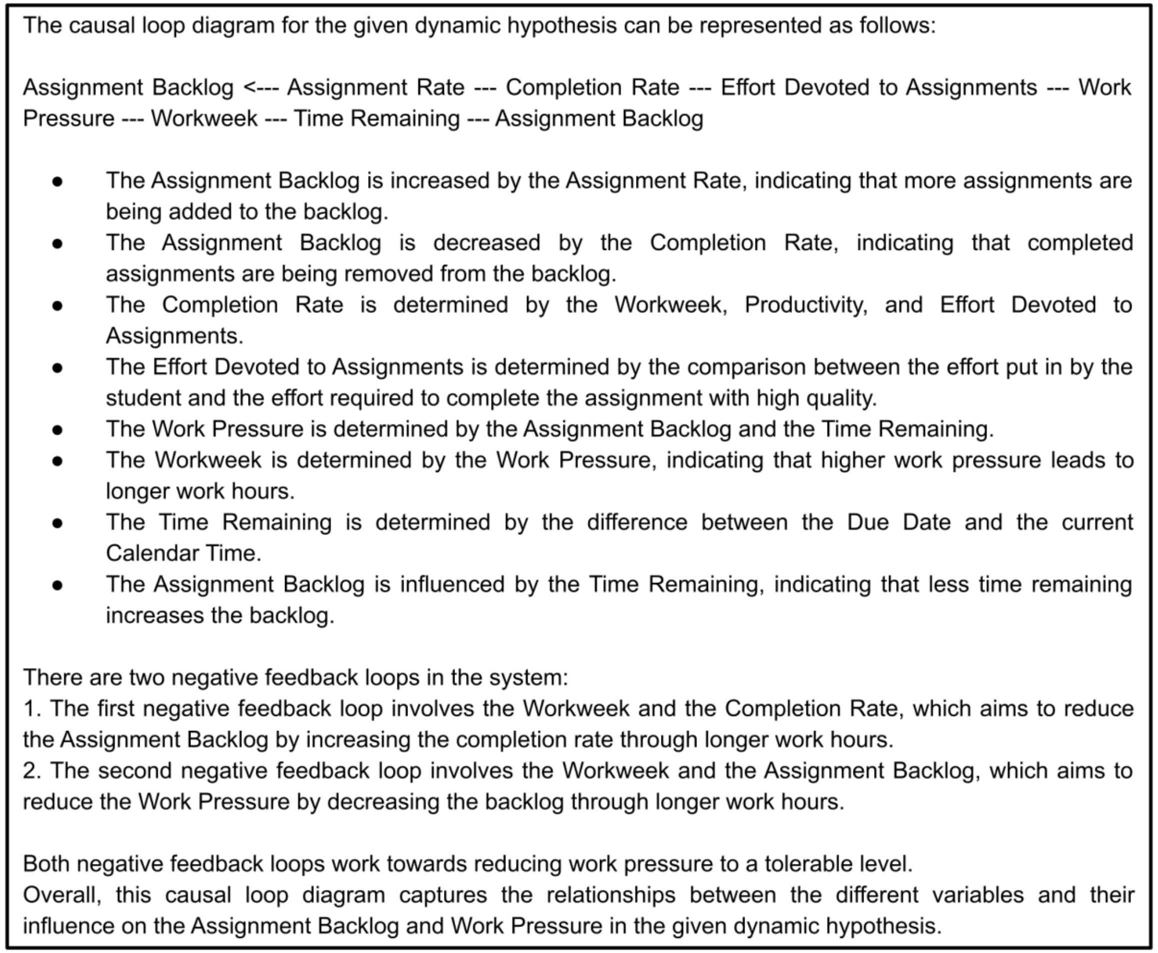
INPUT: “Car production builds the inventory of cars at the dealer. A **higher inventory** leads to a **lower market price**, and **lower market prices** cause **less car production** in the future. If the **price** were to **increase**, the **retail sale of cars** would tend to **fall**. **Retails sales** drain the **inventory of cars** held in stock at the dealership. And **a decline in the inventory** will cause the dealers to **raise** their **prices** in the future.”

LLM Generated Outputs			Ground truth (From SD textbook)
Few-shots Approach	Guided Prompts Approach	Two-stage Approach	Expert CLD
 Retail Sales Car Production Inventory Price Market Price	 car production inventory market price price retail sales	 inventory of cars at dealership market price car production retail car sales	 inventory of cars at dealership market price car production retail car sales

Note: “->” indicates a positive relationship, “-|” indicates a negative relationship

Case 3: Two Balancing Loop with Exogenous Variables - Assignment Backlog

INPUT: *"The **Assignment Backlog** is increased by the Assignment Rate and decreased by the Completion Rate. Completion Rate is Workweek (hours per week) times Productivity (tasks completed per hour of effort) times the Effort Devoted to Assignments. Effort Devoted to Assignments is the effort put in by the student compared to the effort required to complete the assignment with high quality. If **work pressure** is high, the student may choose to **cut corners,** skim some reading, skip classes, or give less complete answers to the questions in assignments. For example, if a student works 50 hours per week and can do one task per hour with high quality but only does half the work each assignment requires for a good job, then the completion rate would be (50)(1)(.5) = 25 task equivalents per week. **Work Pressure determines the workweek and effort devoted to assignments. Work pressure** depends on the **assignment backlog** and the **Time Remaining** to complete the work: The bigger the **backlog** or the less time remaining, the higher the workweek needs to be to complete the work on time. Time remaining is of course simply the difference between the Due Date and the current Calendar Time. The two most basic options available to a student faced with **high work pressure** are to first, **work longer hours**, thus increasing the completion rate and reducing the backlog , or second, **work faster by** spending less time on each task, speeding the completion rate and reducing the backlog. Both are **negative feedbacks** whose goal is to reduce work pressure to a tolerable level." (Sterman, 2000)*



The causal loop diagram for the given dynamic hypothesis can be represented as follows:

Assignment Backlog <--- Assignment Rate --- Completion Rate --- Effort Devoted to Assignments --- Work Pressure --- Workweek --- Time Remaining --- Assignment Backlog

- The Assignment Backlog is increased by the Assignment Rate, indicating that more assignments are being added to the backlog.
- The Assignment Backlog is decreased by the Completion Rate, indicating that completed assignments are being removed from the backlog.
- The Completion Rate is determined by the Workweek, Productivity, and Effort Devoted to Assignments.
- The Effort Devoted to Assignments is determined by the comparison between the effort put in by the student and the effort required to complete the assignment with high quality.
- The Work Pressure is determined by the Assignment Backlog and the Time Remaining.
- The Workweek is determined by the Work Pressure, indicating that higher work pressure leads to longer work hours.
- The Time Remaining is determined by the difference between the Due Date and the current Calendar Time.
- The Assignment Backlog is influenced by the Time Remaining, indicating that less time remaining increases the backlog.

There are two negative feedback loops in the system:
1. The first negative feedback loop involves the Workweek and the Completion Rate, which aims to reduce the Assignment Backlog by increasing the completion rate through longer work hours.
2. The second negative feedback loop involves the Workweek and the Assignment Backlog, which aims to reduce the Work Pressure by decreasing the backlog through longer work hours.

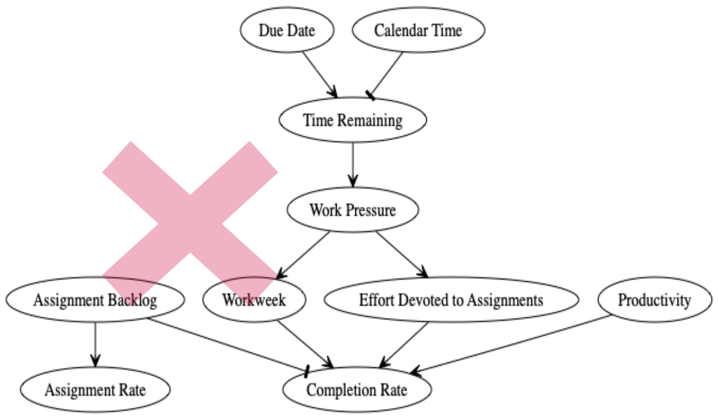
Both negative feedback loops work towards reducing work pressure to a tolerable level.
Overall, this causal loop diagram captures the relationships between the different variables and their influence on the Assignment Backlog and Work Pressure in the given dynamic hypothesis.

Figure 1. Generated Outputs from Approach 1

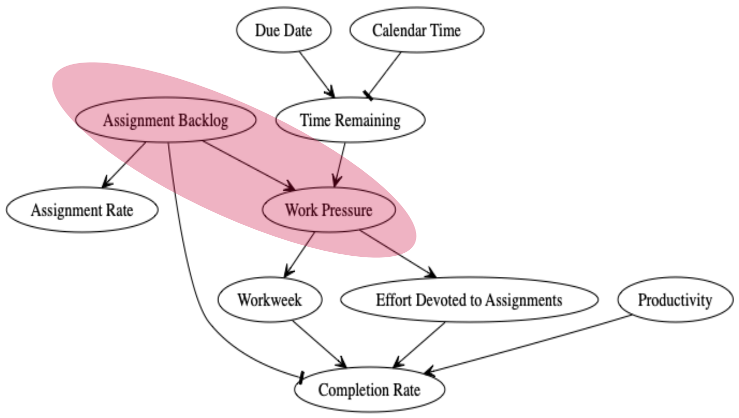
Case 3: Two Balancing Loop with Exogenous Variables - Assignment Backlog

LLM Generated Outputs

Few-shots Approach

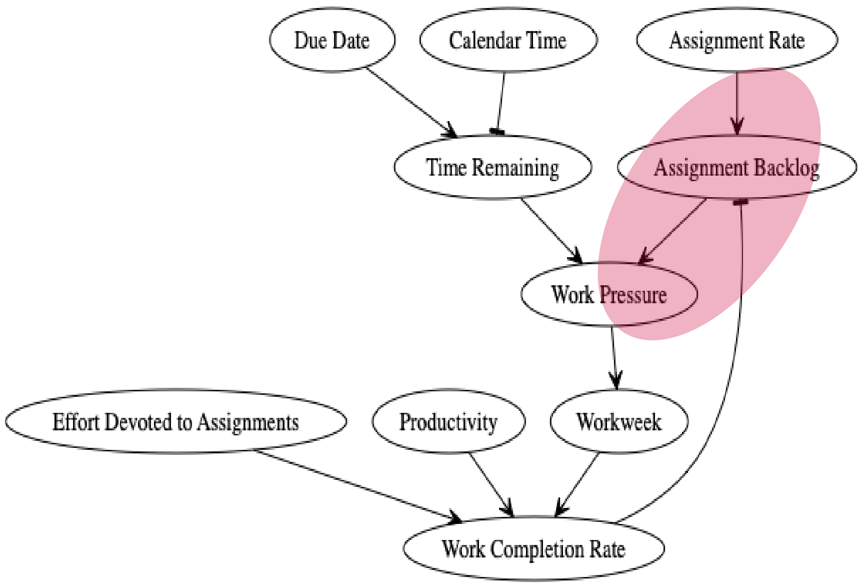


Guided Prompts Approach



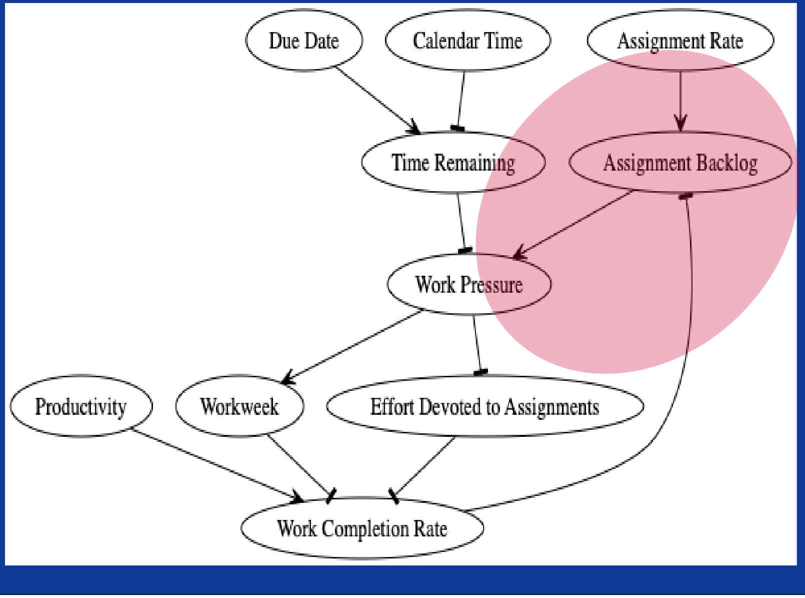
LLM Generated Outputs

Two-stage Approach



Ground truth (From SD textbook)

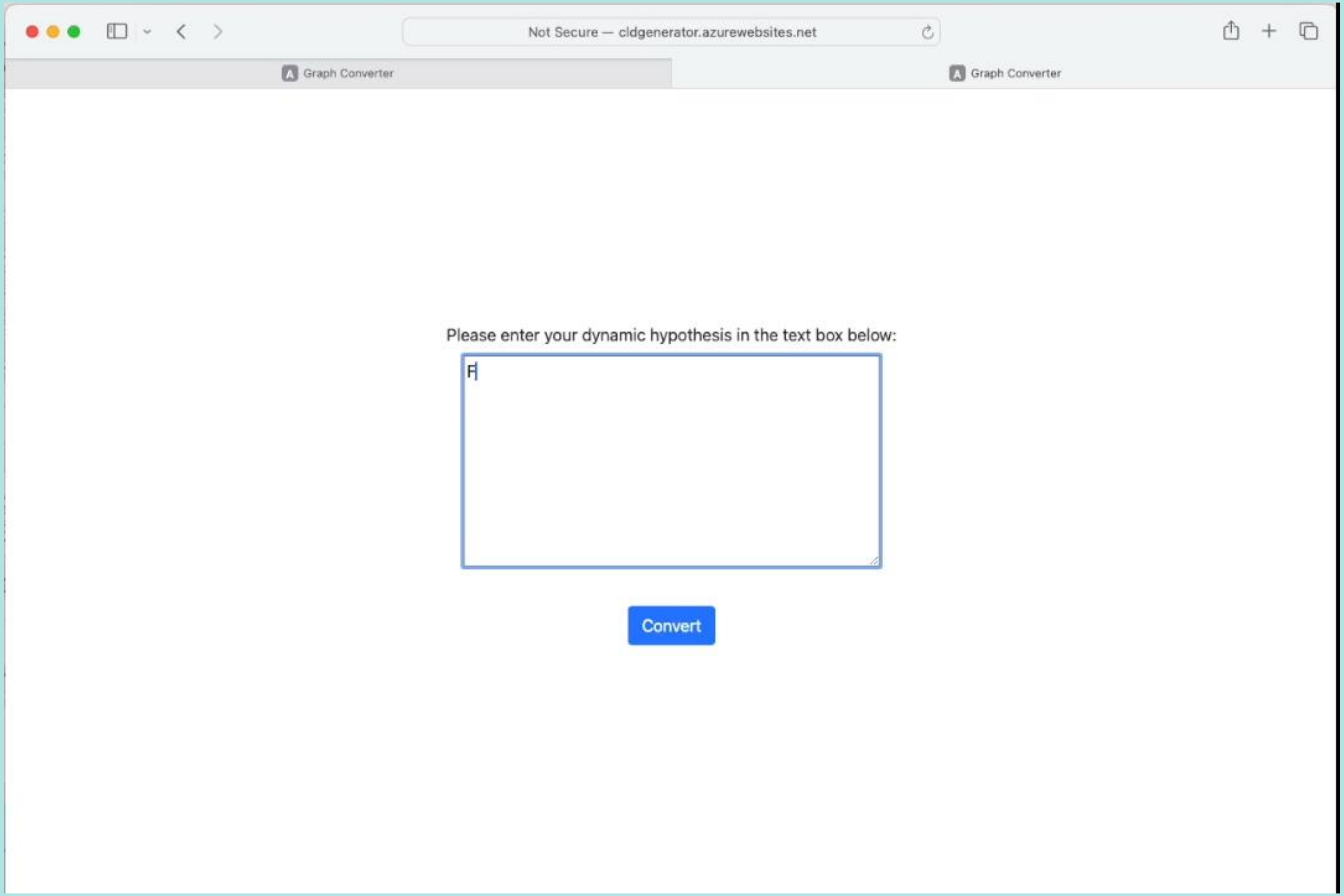
Expert CLD



Note:
“->” indicates a positive relationship, “**-|**” indicates a negative relationship

CLD Generator Demo

SCAN ME



Please enter your dynamic hypothesis in the text box below:

F

Link to the website: http://cldgenerator.azurewebsites.net

Conclusions

Lesson Learnt, Limitations and Future research

Conclusion

SCAN ME

Demonstrated the potential of LLMs to accelerate the analysis of complex systems:

- ✓ LLMs can generate CLDs comparable to expert human modelers for simple feedback structures.
- ✓ Curated prompting techniques improve the quality of generated CLDs.

Benefits of integrating LLMs into the System Dynamics (SD) Modelling toolkit:

- ✓ Accelerates CLD development process.
- ✓ Lowers barriers for novice modelers.
- ✓ Aids growth of the SD field.
- ✓ Enhances quality standards in SD modeling.

Limitations and Future Research

Establish Standardized Measurement of CLDs for Effective Benchmarking:

- Developing consistent metrics for evaluating Causal Loop Diagrams (CLDs) is crucial to scalability and to enable comparative analysis and benchmarking across different datasets and studies.

Expand Future Research on Interview Transcripts:

- To enhance practicality, future studies should focus on analyzing unstructured data sources, such as interview transcripts, to capture real-world complexities better.

Create a Centralized Repository for Textual Data<> CLDs:

- Building a comprehensive data repository for qualitative data will facilitate broader access to valuable insights and support the development of robust CLDs through diverse datasets.

Recap

Failure to Understand Complex Systems Can Lead to Disasters:

- Understanding complex systems is essential.
- System Thinking (ST) and Casual Loop Diagram (CLD) are ways to represent one's understanding of a complex system.

How Can We Better Measure One's ST Skill Considering the Multi-Dimensionality Characteristic?

- A Multidimensional Index of Systems Thinking Skills From Textual Data (Paper 1).

How Can We Improve the Manual Process of Translating Textual Data into CLDs?

- Leveraging Large Language Models for Automated Causal Loop Diagram Generation with Curated Prompting (Paper 2) .

Thank you for your attention!

Q&A

Georgia Ning-Yuan Liu

Contact me at: gliu26@mgh.harvard.edu

Personal Website: https://georgia-max.github.io/