

Praxis Development for Artificial Intelligence

Professor Hamza F. Alsarhan

SEAS 8599 – DA1

Lecture 2

April 27, 2024



Any Questions?

- Any questions about HW #1?
- Any questions about lecture 1?
- Any other questions?

Agenda

1. Problem Statement
2. Thesis Statement
3. Research Questions
4. Research Hypotheses

Problem Statement

1. Definition
2. Formulation
3. Checklist
4. Elaboration
5. Examples

Problem Statement – Definition

- A problem statement is the first step in research formulation, and it addresses the following questions:
 - What is the problem at hand?
 - Why is this problem important?
- The problem statement captures the issue that you plan to research and its importance (the “so what”).
- The structure of a problem statement is:
 - Issue: What is the problem being addressed?
 - So What: Why is this problem important?

Problem Statement – Formulation

- Single clear and concise sentence (i.e. , 30 words or less)
 - Aim to make your problem statement easy to understand.
 - It does not require expertise in a particular field to be understandable.
 - Avoids acronyms and jargon.
 - Long sentences or a paragraph muddle the idea and lose the reader.
- Uses specific language
 - A problem statement should not use vague terms, such as “most” or “best”.

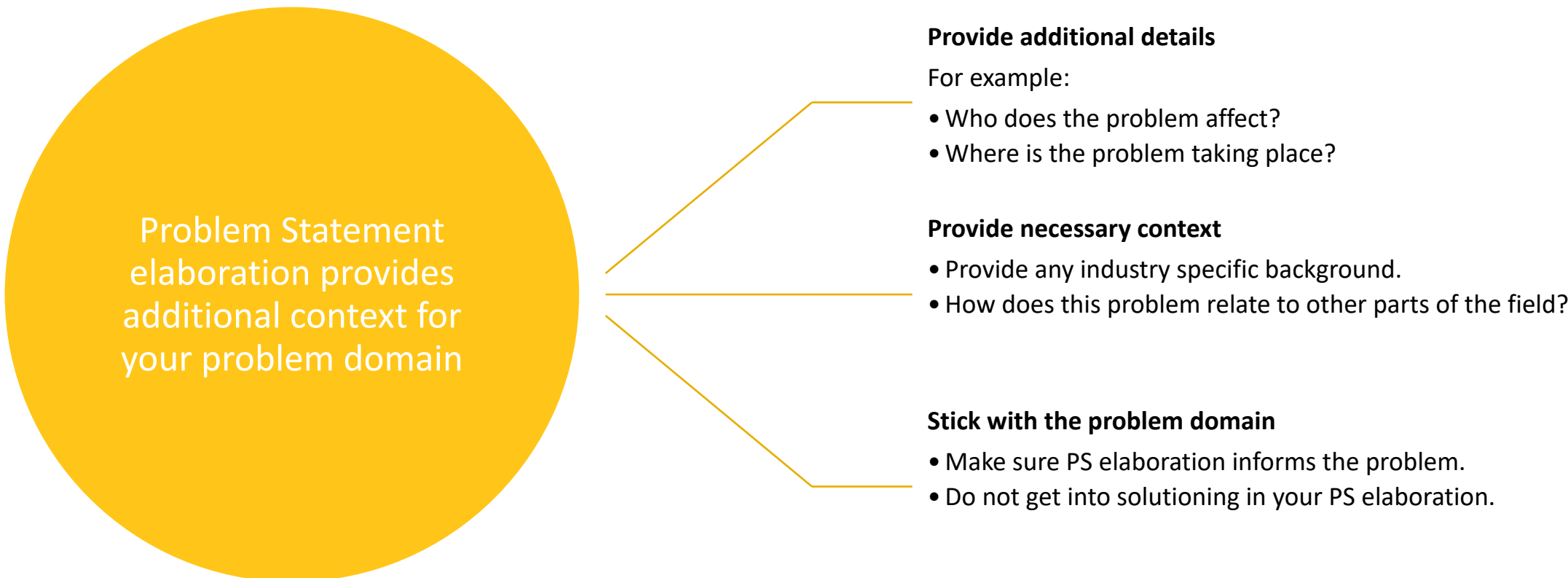
Problem Statement – Formulation

- Discusses an issue that has occurred or is occurring
 - Do not focus on potential issues that may occur in the future.
 - Focus on issues that are relevant.
- You can always refine your problem statement as you go deeper into your research
 - Do not expect your problem statement to be finalized until you are completely done with the research phase.

Problem Statement – Checklist

- **Single Issue** – focuses on a single issue and does not attempt to address various topics.
- **Concise** – has less than 30 words.
- **Not too broad** – stays specific to the clear topic.
- **Not hinting at a solution** – focuses on the problem. For example, do not use the word “lack” in your problem statement. If there is a lack of something, then, in essence, a solution is being proposed.
- **Has structure** – begins with an issue and ends with a “so what”. Not the other way around.

Problem Statement – Elaboration



Problem Statement
elaboration provides
additional context for
your problem domain

Provide additional details

For example:

- Who does the problem affect?
- Where is the problem taking place?

Provide necessary context

- Provide any industry specific background.
- How does this problem relate to other parts of the field?

Stick with the problem domain

- Make sure PS elaboration informs the problem.
- Do not get into solutioning in your PS elaboration.

Problem Statement – Examples

Example 1

- Tanks aboard U.S. Navy surface ships are being inspected too frequently, wasting money and maintenance time.
- Comment on the above PS.

Example 2

- The cloud security solutions are overwhelmingly broad and complex, which presents significant challenges for IT managers to prioritize the solutions and allocate the limited budget efficiently, hence leaving organizations' cloud environments vulnerable to data breaches.
- Comment on the above PS.

Problem Statement – Examples

Example 1 Review

- Tanks aboard U.S. Navy surface ships are being inspected too frequently, wasting money and maintenance time.
 - **Issue:** Tanks aboard U.S. Navy Surface ships are being inspected too frequently
 - **“so what”:** Wastes money and maintenance time
 - “so what” is not specific. How much money and maintenance time is being wasted is not clear. If the amounts are too little then perhaps it is not worth studying but if amounts are substantial then it is a legitimate problem.

Example 2 Review

- The cloud security solutions are overwhelmingly broad and complex, which presents significant challenges for IT managers to prioritize the solutions and allocate the limited budget efficiently, hence leaving organizations’ cloud environments vulnerable to data breaches.
 - **Issue:** The cloud security solutions are overwhelmingly broad and complex, which presents significant challenges for IT managers to prioritize the solutions and allocate the limited budget efficiently
 - **“so what”:** leaves organizations’ cloud environments vulnerable to data breaches
 - PS is chatty. It is not < 30 words.
 - A number of vague words (i.e. overwhelmingly, broad, complex, significant, challenge)
 - “so what” is not specific.

Problem Statement – Examples

Example 3

- Electric utility investors cannot predict total capital costs or life cycle costs of energy storage system installations needed to calculate return on investment.
- Comment on the above PS.

Example 4

- Low-volume medically significant products are subject to miss delivery dates and contribute to supply disruptions.
- Comment on the above PS.

Problem Statement – Examples

Example 3 Review

- Electric utility investors cannot predict total capital costs or life cycle costs of energy storage system installations needed to calculate return on investment.
- **Issue:** Electric utility investors cannot predict total capital costs or life cycle costs of energy storage system installations needed to calculate return on investment.
- **"So what":** ?
- The issue is written poorly. Better way to capture the issue: "Electric utility investors are unable to calculate return on investment for energy storage system installations"
- "So what" is missing!!

Example 4 Review

- Low-volume medically significant products are subject to miss delivery dates and contribute to supply disruptions.
- **Issue:** Low-volume medically significant products are subject to miss delivery dates and contribute to supply disruptions
- **"So what":** ?
- Not a strong issue statement. "are subject to" implies that it could be a problem, or it could not.
- The word "significant" is vague.
- "So what" is missing.

Problem Statement – Examples

Example 5

- The shortage of electric power within Togo's electric power system, which is estimated to be 496 GWh per year on average, contributes to lower-than-expected GDP.
- Comment on the above PS.

Example 6

- Lack of a framework to properly select Agile over other project development methodologies can decrease an organization's average project success rate by 9%.
- Comment on the above PS.

Problem Statement – Examples

Example 5 Review

- The shortage of electric power within Togo's electric power system, which is estimated to be 496 GWh per year on average, contributes to lower-than-expected GDP.
- **Issue:** The shortage of electric power within Togo's electric power system, which is estimated to be 496 GWh per year on average
- **"So what":** lower than expected GDP.
- Issue can be stated more succinct: "Togo electric power system fails to provide approximately 500 GWh of demand".
- "So what" is vague. "lower than expected" is not specific. Need to state in terms of percentage or dollar.
- One thing to keep an eye on is scope of this problem.

Example 6 Review

- Lack of a framework to properly select Agile over other project development methodologies can decrease an organization's average project success rate by 9%.
- **Issue:** Lack of a framework to properly select Agile over other project development methodologies.
- **"so what":** decrease an organization's average project success rate by 9%.
- Issue: The PS is hinting at a solution!! Need to rewrite to discuss the real underlying issue.

Problem Statement – Examples

Example 7

- On-site heating processes used for oil shale extraction continue to be underutilized due to cost uncertainties, leading to a potential loss of \$90 billion per year in revenue.
- Comment on the above PS.

Example 8

- Defects in the output of an aerospace batch manufacturing process has resulted in delayed deliveries and loss of revenue exceeding 10%.
- Comment on the above PS.

Problem Statement – Examples

Example 7 Review

- On-site heating processes used for oil shale extraction continue to be underutilized due to cost uncertainties, leading to a potential loss of \$90 billion per year in revenue.
 - **Issue:** On-site heating processes used for oil shale extraction continue to be underutilized due to cost uncertainties
 - **“So what”:** leading to a potential loss of \$90 billion per year in revenue
 - “Cost Uncertainties” is a bit vague, but other than that, it looks good.

Example 8 Review

- Defects in the output of an aerospace batch manufacturing process has resulted in delayed deliveries and loss of revenue exceeding 10%.
 - **Issue:** Defects in the output of an aerospace batch manufacturing process
 - **“So what”:** delayed deliveries and loss of revenue exceeding 10%
 - Delayed deliverables should be quantified.
 - Need to quantify the loss of revenue in terms of \$s.

Problem Statement – Examples

Example 9

- Large Internet companies are unable to estimate the ROI of open-sourcing their software resulting in higher maintenance costs in up to 80% of software projects.
- Comment on the above PS.

Example 10

- The current Department of Defense process for integrating Rapid Response Projects into Large Weapon Systems Programs results in a less than 45% success rate.
- Comment on the above PS.

Problem Statement – Examples

Example 9 Review

- Large Internet companies are unable to estimate the ROI of open-sourcing their software resulting in higher maintenance costs in up to 80% of software projects:
 - **Issue:** Large Internet companies are unable to estimate the ROI of open-sourcing their software
 - **“So what”:** higher maintenance costs in up to 80% of software projects
 - Overall, it is okay.

Example 10 Review

- The current Department of Defense process for integrating Rapid Response Projects into Large Weapon Systems Programs results in a less than 45% success rate.
 - **Issue:** The current Department of Defense process for integrating Rapid Response Projects into Large Weapon Systems Programs
 - **“So what”:** less than 45% success rate
 - Needs a more precise “so what”. The actual problem is not clear in the issue. What is the issue with the current process? Specifically state that in the issue part of the PS.

Problem Statement – Examples

Example 11

- Lack of clear guidelines on selecting risk identification techniques during the NASA project life cycle can negatively impact project performance, contributing to cost growth of 27.6% and 13 months (average) launch delays.
- Comment on the above PS.

Example 12

- AWS instance upgrade in USCIS leaves some applications with too little capacity while others remain underutilized, resulting in a year-over-year 30% growth in instance upgrade costs.
- Comment on the above PS.

Problem Statement – Examples

Example 11 Review

- Lack of clear guidelines on selecting risk identification techniques during the NASA project life cycle can negatively impact project performance, contributing to cost growth of 27.6% and 13 months (average) launch delays.
- **Issue:** Lack of clear guidelines on selecting risk identification techniques during the NASA project life cycle
- **“So what”:** negatively impact project performance, contributing to cost growth of 27.6% and 13 months (average) launch delays
- The above PS is hinting at a solution! Do not state your issue in terms of “lack of something”.

Example 12 Review

- AWS instance upgrade in USCIS leaves some applications with too little capacity while others remain underutilized, resulting in a year-over-year 30% growth in instance upgrade costs.
- **Issue:** AWS instance upgrade in USCIS leaves some applications with too little capacity while others remain underutilized
- **“So what”:** resulting in a year-over-year 30% growth in instance upgrade costs.
- All is good!

Problem Statement – Examples

Example 13

- Agile practitioners fail to write good User Stories, which negatively impacts system quality and contributes to 13% of canceled projects.
 - Comment on the above PS.

Example 14

- Network latencies are degrading Connected Vehicle Systems' safety, preventing meeting the US Department of Transportation mandated 20% reduction in US crash fatalities by 2020 and beyond.
 - Comment on the above PS.

Problem Statement – Examples

Example 13 Review

- Agile practitioners fail to write good User Stories, which negatively impacts system quality and contributes to 13% of canceled projects.
 - **Issue:** Agile practitioners fail to write good User Stories
 - **"So what":** negatively impacts system quality and contributes to 13% of canceled projects.
 - Word "good" is vague.

Example 14 Review

- Network latencies are degrading Connected Vehicle Systems' safety, preventing meeting the US Department of Transportation mandated 20% reduction in US crash fatalities by 2020 and beyond.
 - **Issue:** Network latencies are degrading Connected Vehicle Systems' safety
 - **"So what":** preventing meeting the US Department of Transportation mandated 20% reduction in US crash fatalities by 2020 and beyond.
 - Overall, it is good!!

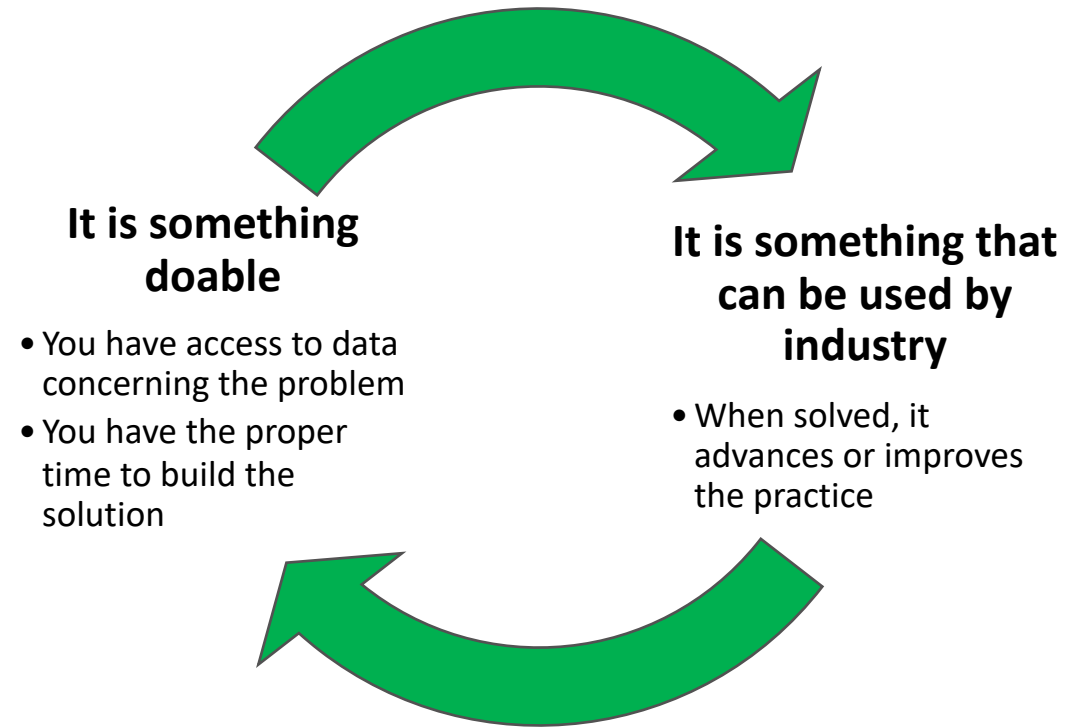
Thesis Statement

1. Definition
2. Formulation
3. Checklist
4. Examples

Thesis Statement – Definition

- Once you have conducted an extensive literature search on the problem domain, you can develop a solution that is missing from existing practice in the industry.
- Thesis statement captures your ***solution to the problem statement***.
 - Should tie back to your problem statement.
- It is your ***claim***
 - It is not a statement of fact. It should be a statement that can be debated.
 - What you put forward for consideration which you will prove in your paper.
 - It should try to convince the reader of your solution.
 - It is the product of your own critical thinking after you have done some research.
 - It is not just your opinion alone!
- It clearly ***states the deliverable*** of your research
 - Make sure the deliverable is clear
 - Example: A predicative model is required to....
 - Example: A Decision Support Tool is required to....

Thesis Statement – Definition



Thesis Statement – Formulation

- Single clear and concise sentence (i.e. 30 words or less)
 - Aim to make your thesis statement easy to understand.
 - Does not require expertise in a particular field to be understandable.
 - Avoids acronyms and jargon.
 - Long sentences or a paragraph muddle the idea and lose the reader.
- Uses specific language
 - A thesis statement should not use vague terms, such as “most” or “best”.

Thesis Statement – Formulation

- Explicitly states the deliverable of the research
 - The Thesis Statement needs to contain a clear deliverable.
 - For example: a decision support tool to reduce..., or a predictive model to...
- You can always refine your thesis statement as you go deeper into your research
 - Do not expect your thesis statement to be finalized until you are completely done with the research phase.

Thesis Statement – Checklist

- **Clear** – states your claim and has a clear deliverable.
- **Concise** – has less than 30 words.
- **Directly links back to your problem statement** – ties back to the PS.
- **Does not start with a verb** – starts with the deliverable (a noun).
- **No personal pronouns** – does not include “I” or “We”.

Thesis Statement – Checklist

- **Certain** – for example, do not say “a predictive model can be developed to solve...”; instead, say “a predictive model will solve...”.
- **Demonstrable** – supported by examples and details throughout the praxis to back up your claim.
- **Arguable** – does not state a fact such as “A machine learning-based model can provide predictions”. Instead provide an arguable claim that can be supported by data, such as “A machine learning-based model will improve the cyber resilience of IoT networks”.

Thesis Statement – Examples

Example 1

- Developing a cost-benefit model based on open-source project data may help significantly reduce evaluation effort and increase the probability of making the optimal decision.
- Comment on the above TS.

Example 2:

- Linear programming models utilizing data from enterprise tools can improve the allocation of DoD resources for cybersecurity readiness.
- Comment on the above TS.

Thesis Statement – Examples

Example 1 Review

- Developing a cost-benefit model based on open-source project data may help significantly reduce evaluation effort and increase the probability of making the optimal decision.
 - Deliverable: Cost-benefit model
 - Starts with a verb!
 - It does not present a claim: “may help”

Example 2 Review

- Linear programming models utilizing data from enterprise tools can improve the allocation of DoD resources for cybersecurity readiness.
 - Deliverable: Linear Programming Models
 - Uses passive voice : “can”

Thesis Statement – Examples

Example 3

- In order to improve the quality of software, more unit testing is required prior to moving the code to acceptance.
- Comment on the above TS.

Example 4:

- An Agile Success Estimation Framework for Software Projects supplements an existing Agile Suitability Filter (PMI, 2017) with a quantified estimate of project success likelihood and potential cost, schedule, quality and stakeholder satisfaction success outcomes achievable by selecting an Agile methodology.
- Comment on the above TS.

Thesis Statement – Examples

Example 3 Review

- In order to improve the quality of software, more unit testing is required prior to moving the code to acceptance.
- Deliverable: ?
- There is nothing arguable in above since it is almost given that more testing would improve quality

Example 4 Review

- An Agile Success Estimation Framework for Software Projects supplements an existing Agile Suitability Filter (PMI, 2017) with a quantified estimate of project success likelihood and potential cost, schedule, quality and stakeholder satisfaction success outcomes achievable by selecting an Agile methodology.
- Deliverable: Agile Success Estimation Framework
- It is not concise! (> 30 words)
- It is vague!

Thesis Statement – Examples

Example 5

- Virginia electric distribution cooperatives need a quantitative model to evaluate the wholesale power cost of lithium-ion utility energy storage systems.
- Comment on the above TS.

Example 6

- A predictive model for prioritizing open-source software vulnerabilities is required to prevent potential data breaches.
- Comment on the above TS.

Thesis Statement – Examples

Example 5 Review

- Virginia electric distribution cooperatives need a quantitative model to evaluate the wholesale power cost of lithium-ion utility energy storage systems.
 - ***Deliverable:*** Quantitative Model
 - It can be improved by being specific (i.e. stating the exact type of quantitative model)

Example 6 Review

- A predictive model for prioritizing open-source software vulnerabilities is required to prevent potential data breaches.
 - ***Deliverable:*** Predictive Model
 - It can be improved by stating that it can help prevent as opposed to prevent. Prioritization of vulnerabilities can not prevent potential data breaches, but it can help prevent potential data breaches.

Thesis Statement – Good Examples

1. A linear programming model utilizing data from enterprise tools is needed to improve the allocation of DoD resources for cybersecurity initiatives.
2. A cost-benefit model based on open-source project data will reduce evaluation efforts of new projects for midsize software development companies.
3. A machine learning-based intrusion detection system (IDS) is required to reduce the impact of cyberattacks on healthcare organizations in the United States.

Research Questions

1. Definition
2. Characteristics
3. Types
4. Examples

Research Questions – Definition

Topic:

- Title of your research.
- Represent a broad area.
- A research question guides and centers your research.

Research Question

- Narrows down your topic.
- Deals with specific aspects of your topic.
- Pick aspects of the topic you want to focus on.

Research Questions – Definition



Frames Your Research

They are a set of questions you like to explore at the core of your research

A set of questions that your research study sets out to answer

Answering your research questions is the main aim of your research



Brings Focus To Research

They are important since they help focus your research

They give direction to your research

Think through what you want to add to the field as you formulate your questions.

What is it that I want to know?



Drive the Design of Research

What data is collected

Data collection methods

Data analysis methods



They are questions

They are not claims or conclusions about an idea

Developing your research questions is the initial step in your research

Research Questions – Characteristics

- **Clear:** measures a single goal with defined parameters. For example: “Are young kids better at geometry than adults?” → “Do kids ages 10-15 score higher than adults ages 30-35 on a standardized geometry exam?”
- **Specific:** defines what is being compared or studied. Stay away from words such as “best”.
- **Answerable:** can be answered through research. Must be either confirmed or negated by your praxis results.

Types of Research Questions



Types of Research Questions – Descriptive

- The goal of descriptive question is:
 - to describe the concept or topic in question.
 - to describe the data.
- What are the characteristics of X?
- Also aims to quantify the variables in your research.
- We use descriptive questions to understand the usage of a product like the frequency, time of day, the purpose of use, etc.
- Common descriptive research questions will begin with “How much?”, “How regularly?”, “What percentage?”, “What time?”, “What is?”

Types of Research Questions – Descriptive

- Examples of Descriptive Research Questions:
 - What is the average number of phone calls to Citibank help desk on a given day?
 - Why does texting influence driving?
 - How many times does a teenager text on average while driving X miles?
 - How frequently do malware attacks occur?
- The researcher cannot infer any conclusions from this type of analysis.
- Descriptive questions do not have corresponding hypotheses because the researcher is not making inferences.

Types of Research Questions – Comparative

- Examines the differences between two or more groups in regard to a variable.
- The questions often begin with “What is the difference in...” or “Is there a difference...”
- What are the differences and similarities between X and Y?
- Examples:
 - What is the difference in number of defects in Release 1.0 vs Release 2.0 of XYZ software?
 - What is the difference in credit score between adults in state A and adults in state B?

Types of Research Questions – Relational

- Explores the relationship, association, trend, or interaction between one or more dependent (e.g., exam scores) and independent (e.g., study time) variables for a given population
- Focus on the existence of a relationship
 - Are variables 1 and 2 related?
 - What is the relationship between variable X and variable Y?
- Focus on cause and effect
 - Does variable 1 cause variable 2?
- Examples:
 - What is the relationship between testing time and the number of defects discovered in a financial software?

Research Questions – Examples

Example 1

- Which machine learning model is the best to predict DDoS attacks on hospital networks in the US?
- Comment on the above RQ.

Example 2

- Is Agile better than waterfall?
- Comment on the above RQ.

Research Questions – Examples

Example 1 Review

- Which machine learning model is the best to detect DDoS attacks on hospital networks in the US?
- Not a good research question.
- The word “best” makes it hard to exhaustively and fully answer the question.
- It is unclear what aspects of a model are being measured.
- ***Revised question: Which of Random Forest and Support Vector Machines can most accurately detect DDoS attacks on hospital networks in the US with the lowest false alarm rate?***

Example 2 Review

- Is Agile better than waterfall?
- Not a good research question.
- The word “better” is unclear. What aspects are you comparing.
- It does not specify what type of projects are being addressed.
- ***Revised question: What percentage of Agile financial software projects are delivered faster and with a higher quality than waterfall projects?***

Research Hypotheses

1. Definition
2. Characteristics
3. Examples

Research Hypotheses - Definition

- Hypotheses are proposed statements. They are not questions. They are not facts.
- Hypotheses are predictions that are grounded and based on experience and research. They are not simply guesses.
- They are testable predictions on what we expect to happen as a result of our research.

Research Hypotheses - Characteristics

- Statement about the relationship between two or more variables.
- Avoids vague terms such as, “better”, “faster”, etc.
- Makes the quantifiable comparison clear.
- Related to the thesis statement.
- It is not formulated as a question.
- It is testable.

Research Hypotheses - Examples

Example 1

- **Hypothesis:** Students get a better grade on their exam if they get 8 or more hours of sleep the night before.
- **Prediction:** It predicts higher grades based on the number of hours of sleep.
- **Testable:** We can set up experiments and collect data on students' exam grades and the number of hours they slept the night before.

Example 2

- **Hypothesis:** When faced with 4 different brands of cereals, kids will prefer brand C.
- **Prediction:** It predicts selection of cereal brand C when faced with different choices.
- **Testable:** We can set up experiments and collect data on choices among cereal brands.

Research Hypotheses - Examples

Example 3

- **Hypothesis:** Variables X and Y out of the set of variables in the dataset are the best predictors of output Z.
- **Prediction:** It predicts that variables X and Y can best predict output Z.
- **Testable:** We can set up experiments and collect data to verify if these two variables can better predict output Z than any other set of variables in the dataset.

Example 4

- **Hypothesis:** A statistical model can be used to predict the relationship between input I and output O.
- **Prediction:** It predicts that a statistical model can be used to predict the relationship between the input and output.
- **Testable:** We can set up experiments and collect data on whether the model can accomplish that or not.

HW #2

- Using the template provided on Blackboard, provide your problem statement, thesis statement, research questions, and research hypotheses.
- Paste in the Scope of Work slide from HW #1.
- Be sure to follow the guidelines provided in this lecture when writing your problem statement, thesis statement, research questions, and research hypotheses.
- Be sure to color any updates you made to your slides from HW #1 in **red**.
- Be sure to follow the template as is without making any changes to its structure or organization.
- Be sure to name your submitted assignment as follows:
lastName_firstName_HW#2_SEAS_8599_DA1.pptx

Next Steps

- Come to office hours with any questions you may have: Mondays and Tuesdays, 7:00 pm – 8:30 pm ET.
- Work on your HW #2 and submit it by **11:59 pm ET on Thursday**.
- Begin/continue reading up on cybersecurity research topics.
- See you next class!

Thank you!