Appendix: Key Concepts for Data Science Projects

This appendix provides explanations of key concepts and frameworks discussed in the MIT Emerging Talent Program, particularly relevant for Milestone 1: Problem Identification. This document can serve as a quick reference for our team.

1. Divergent and Convergent Thinking

Divergent thinking is a thought process used to generate creative ideas by exploring many possible solutions. It typically occurs in a spontaneous, free-flowing, non-linear manner. In the context of your data science project, divergent thinking is crucial during the initial brainstorming phases, such as identifying potential problems based on personal experiences or exploring various aspects of a domain. It's about opening up possibilities and exploring a wide range of ideas without immediate judgment or constraint.

Convergent thinking, on the other hand, is a thought process that follows a particular set of logical steps to arrive at one correct solution. It is systematic, logical, and often involves narrowing down options. In your project, convergent thinking comes into play when you move from a broad set of ideas to defining a specific problem statement, formulating an actionable research question, or selecting the most appropriate data and methods. It's about making decisions and focusing on a single, well-defined path.

Balance is Key: Your workshops emphasized the importance of balancing these two modes of thinking. You diverge to explore and understand the problem space fully, and then converge to define a clear, actionable problem and research question. This iterative process ensures that you don't prematurely narrow your focus or get lost in too many possibilities.

2. Systems Thinking

Systems Thinking is a holistic approach to analysis that focuses on the way a system's constituent parts interrelate and how systems work over time and within the context of larger systems. Instead of looking at isolated events, systems thinking encourages you to identify patterns, underlying structures, and mental models that contribute to observed phenomena.

As discussed in your workshops, a common metaphor for systems thinking is the **"Iceberg Model"**:

- **Events (Tip of the Iceberg):** These are the visible, immediate occurrences or symptoms of a problem (e.g., a student dropping out of an online course).
- **Patterns/Trends:** By observing multiple events over time, you can identify recurring patterns or trends (e.g., a consistent increase in dropout rates in online courses after the first month).
- **Underlying Structures:** These are the organizational structures, policies, power dynamics, and relationships that create the patterns (e.g., course design flaws, lack of instructor feedback mechanisms, inadequate technical support).
- **Mental Models (Deepest Level):** These are the deeply ingrained assumptions, beliefs, and values that shape how people act and how the system operates (e.g., the belief that online learning is inherently less engaging, or that students should be entirely self-motivated).

By using systems thinking, our team can move beyond addressing mere symptoms to identifying and addressing the root causes of problems, leading to more sustainable and impactful solutions. It helps in gaining a comprehensive understanding of the problem domain.

3. Design Thinking

Design Thinking is a human-centered, iterative process that designers use to solve problems. It is a creative and practical method for problem-solving that focuses on understanding the user's needs and developing innovative solutions. Your workshops highlighted its core principles and stages:

Core Principles:

- **Human-Centered:** Solutions are developed with a deep understanding of the people they are designed for. This involves empathy and putting the user at the center of the process.
- **Iterative:** It's not a linear process. Solutions are continuously refined and improved through cycles of prototyping and testing. If a solution doesn't work, you go back and iterate.
- **Visual:** It often involves visual tools like sketches, diagrams, and post-it notes to facilitate understanding and communication among team members and stakeholders.
- **Collaborative:** It emphasizes teamwork and diverse perspectives, recognizing that the best solutions often emerge from collective intelligence.

Five Stages (as discussed in your workshop, with focus on Milestone 1):

1. **Empathize:** Gain a deep understanding of the users and their needs, challenges, and experiences. This involves observing, engaging, and immersing yourselves in their world. For Milestone 1, this means understanding the problem from the perspective of those affected.
2. **Define:** Clearly articulate the problem you are trying to solve based on your empathy research. This stage involves synthesizing your observations into a clear, human-centered problem statement. This is a key focus for Milestone 1.
3. **Ideate:** Brainstorm a wide range of creative solutions to the defined problem. This is a divergent thinking stage where quantity of ideas is preferred over quality initially.

4. **Prototype:** Build tangible representations of your ideas, even if they are rough or low-fidelity. The goal is to quickly create something that can be tested.

5. **Test:** Put your prototypes in front of users to gather feedback and refine your solutions. This stage often reveals new insights that lead to further iterations.

For Milestone 1, the primary focus is on the **Empathize** and **Define** stages, as these are crucial for framing an actionable research question before moving into solution-oriented phases.

4. Research Question

A **Research Question** is a clear, focused, concise, complex, and arguable question around which you center your research. It is the core of our data science project, guiding all subsequent steps from data collection to analysis and communication of results.

Characteristics of a Good Research Question:

- **Specific:** Avoids ambiguity and clearly defines what is being investigated.
- **Measurable/Answerable:** Can be addressed and answered using data and appropriate methodologies.
- **Achievable:** Feasible to answer within the given constraints (time, resources, data availability).
- **Relevant:** Addresses a significant problem and has potential for impact.
- **Actionable:** The answer to the question should lead to insights that can inform decisions or interventions.

As highlighted in your workshops, a well-formulated research question acts as the “bullseye” of your project, ensuring focus and alignment throughout the entire data science lifecycle. It connects what you are measuring, the problem indicators, and

who is affected by the problem.