Unit 2.1

1. **What is explainability in AI, and why is it important?**
Explainability refers to the ability to understand and interpret the behavior and decisions of AI systems. It is crucial as AI systems become smarter, raising questions about their decision-making processes and ensuring trust and transparency in their operations.

2. What is the difference between white-box and black-box AI systems?

- **White-box AI systems**: Provide access to their internal workings, making it easier to interpret their behavior.
- **Black-box AI systems**: Do not reveal internal processes due to complexity, confidentiality, or intellectual property reasons, making them harder to understand.

3. How can we make AI systems interpretable?
AI systems can be made interpretable by using models like linear regression or decision trees, which allow predictions to be directly traced back to the model's internal structure, enabling an understanding of its behavior.

4. What are post-hoc explanations in AI?
Post-hoc explanations involve interpreting the decisions of an AI model after they are made. These explanations rely on understanding the model's input-output behavior without direct access to its internal parameters.

5. How does the concept of explainability in AI relate to human behavior?
Similar to understanding human decisions by asking for explanations, explainability in AI can involve generating reasons for the AI's decisions, even when direct access to its internal workings isn't available. This approach ensures that decisions are transparent and understandable.

6. What is an explanation in the context of decision-making?
An explanation is any information that sheds light on why a particular decision was made. It involves identifying the factors that influenced the decision, such as carrying an umbrella because it was raining or based on a weather report predicting rain.

7. What are counterfactual explanations?
Counterfactual explanations explore how a decision would change if certain factors were altered. For example, would you still carry an umbrella if it were sunny outside? These explanations help identify what truly drives a decision.

8. How do explanations help in improving AI systems?
Explanations allow developers to understand why AI makes specific decisions. This helps identify flaws, such as relying on irrelevant factors, and improve the system’s generalizability and ethical performance.

9. How do modern devices and AI challenge traditional privacy norms?

Modern devices like smartphones, smartwatches, and IoT devices generate extensive personal data. AI and machine learning can infer additional information from these datasets, raising complex privacy concerns. This unit addresses such challenges, emphasizing the need for compliance with privacy laws and respecting user expectations.

10. How do individuals' perceptions of privacy vary?

The transcript highlights that privacy preferences vary significantly among individuals. Factors like the type of data collected, the method of collection, and its frequency influence comfort levels. For example, while some may agree to self-report driving habits, others may hesitate to share data collected via sensors or GPS.

11. What is meant by 'Privacy by Design' in AI development?

Privacy by Design is a framework that integrates privacy considerations throughout the development lifecycle of AI systems. It includes practices like data minimization, threat modeling, and employing methodologies such as the Linden framework to mitigate privacy risks.

;