

## Outline for XRL tutorial

### 1. Quick recap & motivation (5 minutes)

- **Why is explainability critical in digital finance?**
  - Regulatory compliance (e.g., AI Act, GDPR, financial audits).
  - Trust in AI-driven decisions (credit scoring, algorithmic trading).
  - Debugging and improving RL models.
- **Key XRL techniques** (Brief recap of lecture concepts):
  - Feature importance (SHAP, LIME).
  - Policy visualization (heatmaps, attention maps).
  - Counterfactual analysis & rule-based explanations.

### 2. Group discussion: Choosing an explanation strategy (10 minutes)

- Students **review their problem description** and discuss:
  - What aspects of the RL model should be explained? (Policy, actions, reward function?)
  - Who needs the explanation? (Regulators, investors, end users?)
  - What techniques seem most appropriate?

### 3. Hands-on design phase (25 minutes)

- **Task:** Students outline an explainability solution for their problem, including:
  - **Data & model:** What RL model is used? What data is involved?
  - **Explanation approach:** What XAI/XRL technique(s) will be applied, and why?
  - **User interaction:** How will explanations be presented (graphs, reports, interactive tools)?
- **Teacher feedback:** Teachers will offer feedback to refine the ideas.

### 4. Presentations (15 minutes)

- Each group gives a **2-minute summary** of their explainability solution.
- Quick **Q&A and feedback** from peers and teachers.

### 5. Wrap-up and key takeaways (5 minutes)

- Highlight **common themes and challenges** in XRL for finance.
- Discuss **future directions** and open research questions.