Hello everyone, I'm Jennifer Nguyen and I will be presenting my contributions on our project "Let's Mine Chess! A Testbed for Pattern Mining".

Our project aims to develop a pattern mining framework using chess data to simulate healthcare data analysis challenges. By identifying patterns in chess games and testing the framework's ability to handle incomplete data, we aim to refine methods for early detection of patterns applicable to healthcare scenarios, and potentially on other dataset from different areas in the future.

During Sprint 1, I focused on foundational tasks. I prepared and attended team meetings to align our efforts on developing the pattern mining framework. Additionally, I completed sections 1.3 and 1.4 of the User Requirements Document, defining the project description, objectives, requirements, constraints, assumptions, and risks. I also created a comprehensive data dictionary for our dataset features and added it to our GitHub repository. This was critical for ensuring clarity and consistency in our dataset features, which is essential for accurate pattern mining. I also added the chess pattern recognition data to GitHub which provides baseline patterns crucial for comparing with our chess data. I also spent time upskilling in KNIME, the data analysis platform we are using for this project.

In Sprint 2, I focused on both coordination and technical implementation. I continued to prepare and attend team meetings and proof-read the requirements document for clarity and accuracy. I split our dataset into training and test datasets, crucial for validating our pattern recognition algorithms, and added the code to GitHub. I created a structured folder system and process guide on GitHub to maintain organised project management. I reviewed the code for mapping chess openings to identified patterns and assessed the pattern recognition algorithms, ensuring they were robust and effective. I also updated the data dictionary with features that we had after data cleaning.

In response to feedback from our supervisors, Behnam suggested creating a Confluence space to organize our documentation. I migrated relevant documents and recordings to Confluence, ensuring all team members have access to the latest information.

Looking ahead, in Sprint 3, I am tasked with testing our pattern recognition algorithms on incomplete data. I will use the pm4py library, which was suggested by our stakeholders, to simulate data incompleteness by randomly removing moves and assessing if our algorithms can still accurately identify chess openings.

For Sprint 4, I plan to fine-tune our pattern recognition algorithms based on the findings from Sprint 3, addressing any weaknesses identified during the incomplete data testing and optimizing our methods for better accuracy and robustness. I will also contribute to writing an experience and summary report with results, as outlined in Milestone 4.