Future Possibilities:

Advanced Knowledge Graphs and Entity Linking

- **Knowledge Graph Construction:** Use NER to build and expand knowledge graphs that map relationships between entities extracted from a vast array of documents.
- **Entity Linking:** Combine NER with entity linking to resolve ambiguities and connect extracted entities to specific entries in a knowledge base. This helps tackle domain specific documents and entities.

Automated Content Summarization and Topic Modelling

- **Entity-Based Summarization:** Improve content summarization by focusing on key entities identified through NER. This could be particularly useful in creating executive summaries or abstract generation for academic papers.
- **Topic Modelling:** Use the entities extracted to cluster documents by topic, facilitating thematic research and enabling users to explore content by subject matter.

Intelligent Document Classification and Tagging

- **Automated Tagging:** Automatically tag documents with relevant entities, improving content organization and retrieval in content management systems.
- **Classification:** Enhance document classification by using entity patterns to identify document types, such as contracts, legal briefs, or medical reports.

Personalized User Experience

- **Customized Content Delivery:** Leverage NER to understand user preferences based on the entities they interact with, enabling more personalized content delivery and recommendations.
- **Interactive Chatbots and Virtual Assistants:** Integrate the NER model with AI-driven chatbots to provide more context-aware responses by identifying key entities in user queries.

Compliance and Risk Management

- **Sensitive Data Identification:** Use NER to identify and flag sensitive entities (e.g., personal information, financial data) in documents to ensure compliance with data protection regulations like GDPR.
- **Contract Analysis:** Automate the extraction of critical terms, dates, and parties involved in documents to streamline contract management and risk assessment.

This list provides some topics which could be considered for future projects. These projects could have the potential to improve knowledge management systems. To achieve this, I would strongly recommend improving the NER/ LLM model first. This can be done by finetuning the model on some additional data, specifically data which has a better distribution, improving the model performance across labels. This is something I would have liked to do but did not have the time for due to the unstructured nature of my internship.

Once the model has been improved, the mentioned projects can be considered for development. Improving the model should not take more than 2 weeks of dedicated work, unfortunately I did not have this time after identifying the short comings of the model.