Q1 — Advantages and disadvantages of using additional retrieved info vs letting LLM answer without it

- Advantages
  - Reduced Hallucination – Graph tech provides verified data (Generally)
  - Updated data
  - FAIR Data principles applied
- *Disadvantages*
  - More sources to query data from
  - Specific sources mentioned only- LLMs can query the web for more sources of answers. (Might lead to wrong answers though)
  - If Wikidata/Wikipedia is down then no result would be acquired
  - Provenance not available

Q2 — Advantages and disadvantages of querying on every user question

- *Advantages*
  - Updated data.
  - Only SHACL conformed data available. (If SHACL is used)
  - Hallucination avoided
  - Provenance availble if required
- *Disadvantages*
  - More overhead
  - Labels and Optional queires are more expensive. So it might increase the cost significantly. (Can be avoided if a cache system is available)
  - Non availability of backend causes issues.

Q3 — If coach info were only available as PDFs

- If scanned PDF- then OCR is required
- PDF/UA format allows machine to read it easily
- If the pdf is only source of truth then it will make the system faster as the data is readily avialble and is not dependant on API to retrieve it.
- May leave provenance to be desired.

Q4 — Potential for agents in this process?

- Yes. Agents could make it seemless:
  - Agents can be used to retain memory -short term/ long term
  - Can breakdown compley task into subtask for more accurate answer.
  - Can be set to automatically update if any changes detected.
  - Can be used to check for multiple sources if required as a fail safe.

Q5 — Benefit of a domain data model

- A domain data model helps:
  - Use of FAIR principles to normalise data
  - Support caching, validation, and easier query generation.
  - Enable richer reasoning
  - Provenance.

- Unwanted data from new sources can be filtered using SHACl.