

# Evaluating NL RE tools (Mylopoulos)

- Context: We have developed a tool for extracting requirements from regulations, e.g., extract requirements for a meeting scheduler from a privacy law [Zeni-REJ-2015].
- (Dubious) Assumption: Legalese  $\subseteq$  NL
- Key considerations in evaluating the tool:
  - ✓ Tool is useful if it reduces manual effort substantially without deterioration in the quality of the answer;
  - ✓ (Hence) Tool must find all/most requirements, otherwise a person has to make a full pass over the regulation to find FNs;
  - ✓ There is no gold standard, even experts make mistakes.

# Position statement (Mylopoulos)

- Tools automate NL RE tasks, or assist people in performing NL RE tasks.
- For automation tools, precision is very important in determining the quality of the tool; e.g., for a tool that takes as input a RS and a regulation and tells you whether RS complies with the regulation, a 50% correct response renders the tool useless.
- For assistive tools, reduction of manual effort while maintaining quality is the ultimate evaluation criterion, and then recall is very important.