Time line of Validation of an AI, A, for task, T

Requirements Engineering

1. Decide measures for evaluating correctness in performing T, e.g.,
   global correctness or recall & precision
   based on expected distribution of possible answers

2. Establish human thresholds for chosen measures and the correct answers by having expert humans perform T on a representative input I

Implementation

3. Build A for T

4. Validate A by running A(I) and using its answers to calculate A's measures.
   If A's measures are better than the humans' then A is validated

5. If validated then
   Run A in production while watching for A's changing the real world and T, leading to a change in requirements

6. If not validated then