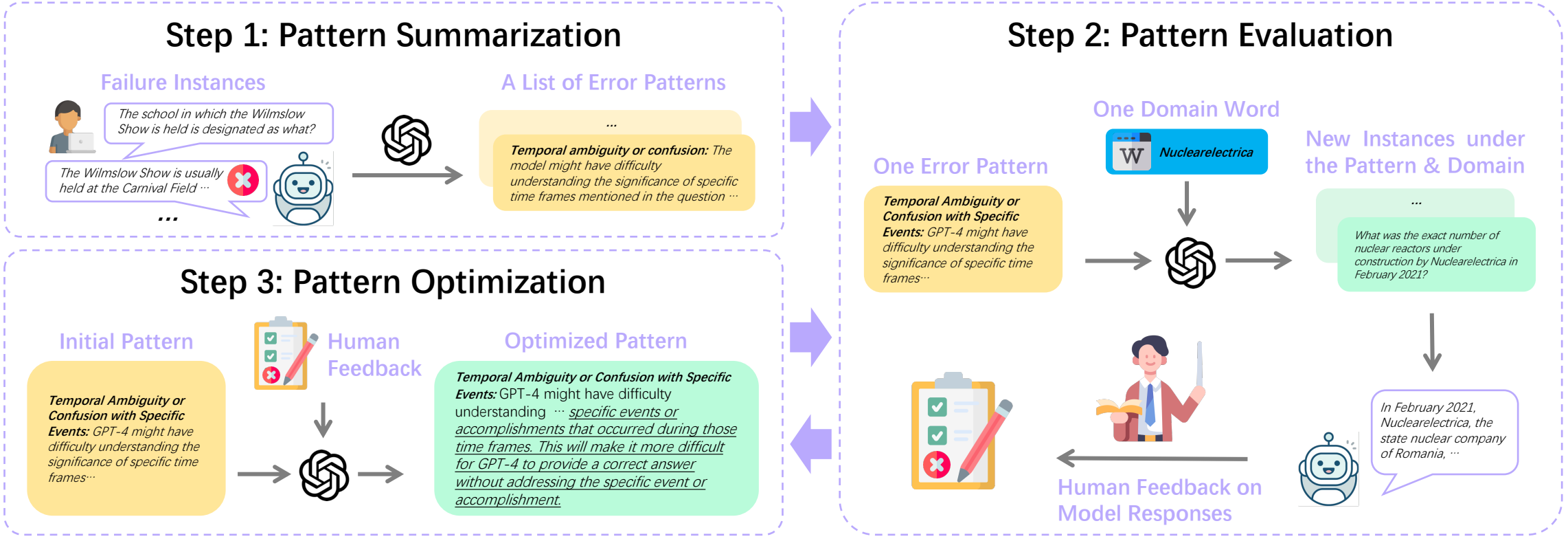

Figure 1: The overall Self-Challenge framework. We first summarize initial error patterns from seed failure instances (Step 1). Then, we perform pattern evaluation (Step 2) to evaluate the quality of summarized patterns, and obtain corresponding human feedback; pattern optimization (Step 3) to modify the original pattern based on human feedback. We frame Step 2 and Step 3 iteratively. We present the difference between Initial Pattern and Optimized Pattern in underline text.