

Strategy: have AI to help humans evaluate multiple reporters

Basic training strategy and why it might work

The strategy involves using an auxiliary AI to help humans improve their understanding of the reports generated by multiple reporters.

The auxiliary AI acts as a mediator that evaluates the outputs from different reporters, helps identify inconsistencies, and provides feedback to both humans and reporters. With that, the reports are not only accurate but also can be understandable to humans. Also, reporters can continually improve based on auxiliary AI feedback.

More precise detail on strategy

While involving humans in the evaluation process, this strategy guarantees that the reports are understandable and useful from a human perspective. The auxiliary AI facilitates better human comprehension. It is responsible to perform cross-validation among reporters: the auxiliary AI compare their outputs and helps identify inconsistencies and inaccuracies. Also, the feedback loop involving humans, the auxiliary AI, and multiple reporters allows for continuous refinement of the reporting process. The auxiliary AI provides targeted feedback to reporters, helping them improve their outputs based on human requirements.

Example:

Considerig a scenario where the SmartVault's reporters are generating reports on potential security threats detected by the sensors, the auxiliary AI compares the outputs from multiple reporters and provides feedback to security personnel. This feedback helps them understand the potential threats and suggests improvements to the reporters. Other methods might fail to ensure comprehensibility or detect subtle inconsistencies, but this strategy guarantees both accuracy and interpretability.