Part 1

Q1. First of all, the patient needs to be informed about the use of an algorithm. Values such as fairness can be further disclosed by the company that owns the algorithms, as well as the translation. The translation transparency has to improve as well, it is unknown what determines the risk and how heavy these things weigh in the decision making process. In addition, when the physicians and policy makers are better informed about potential mistakes like this the consequences of this outcome may not directly result in a refusal to continue treatment. In other words, performance transparency has to improve.

Q2. In my opinion, the physician has the ultimate responsibility for decision making in healthcare. The oath of Hippocrates is sacred in medicine. However, when physicians are obliged to use this, their involvement is minimal. This yields a gap in the accountability of their decisions.

Q3. An additional decision procedure needs to be added into the process in order to bring context into the risk assessment. This cannot be evaluated by a model, so a person needs to get involved. This improves transparency but reduces efficiency.

Q4. The weights of the decision making parameters and the biases involved need to be explainable. This is because in this case the consequences of the decisions can be big for an individual. In addition, the performance of the algorithm is hard to determine for experts when some decision making parameters are unexplained.

Part 2

Q1. Yes I do, when explainability of the model is limited, so may be the understanding of the model among professionals. This can lead to mistakes being made in testing or in training. When mistakes like these are made in early stages like training or testing and a model is delivered with flaws and poor performance, trust in the model from healthcare providers decreases and clinical implementation is stalled.

Q2. The use of models like these for triage decisions comes with a trade off between efficiency and quality. Wrong decisions may be made by an AI model when a patient population is for instance underrepresented in training data.

Q3. The responsibility for the failed implementation of AI models is significant for the developers, since they are responsible for training and testing. They have the most knowledge of the data that is used for these steps and should be able to explain why this data was chosen in the first place. When mistakes are made, like using mislabelled data or data from unknown sources, the developers take a risk of delivering an inadequate model.

Q4. Contestability is of utmost importance in clinical settings, physicians must be able to question and review choices that are made by AI models, and when necessary, override them. However, other requirements should not be overlooked, the use of properly labelled training and testing data is crucial and the source of this data needs to be known by the developers.