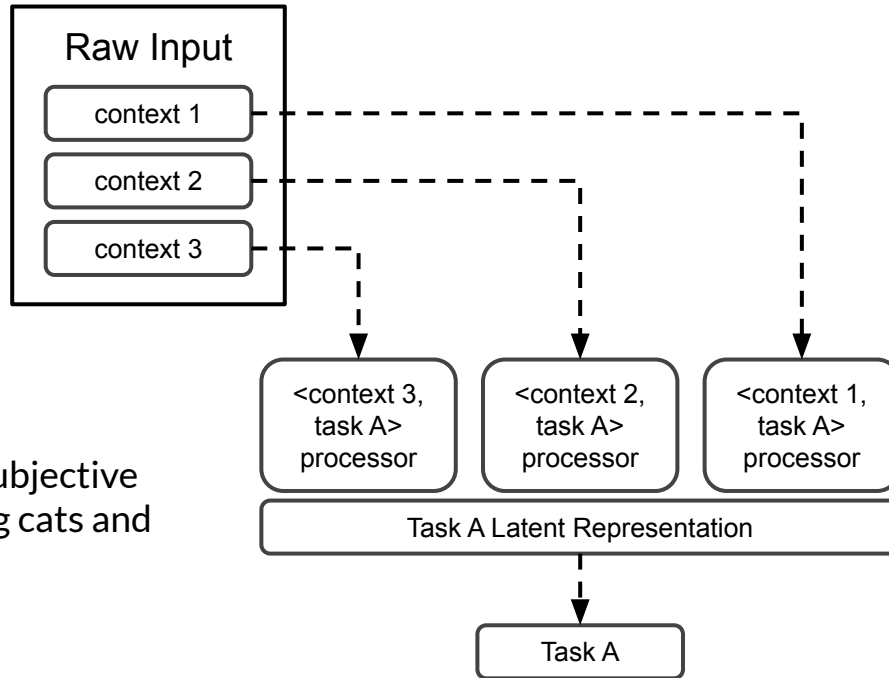
A large red square with a white border, centered on a white background. Inside the square, the text "Meaning-Preserving Continual Learning v1" is written in white.

Meaning-Preserving Continual Learning v1

Meaning

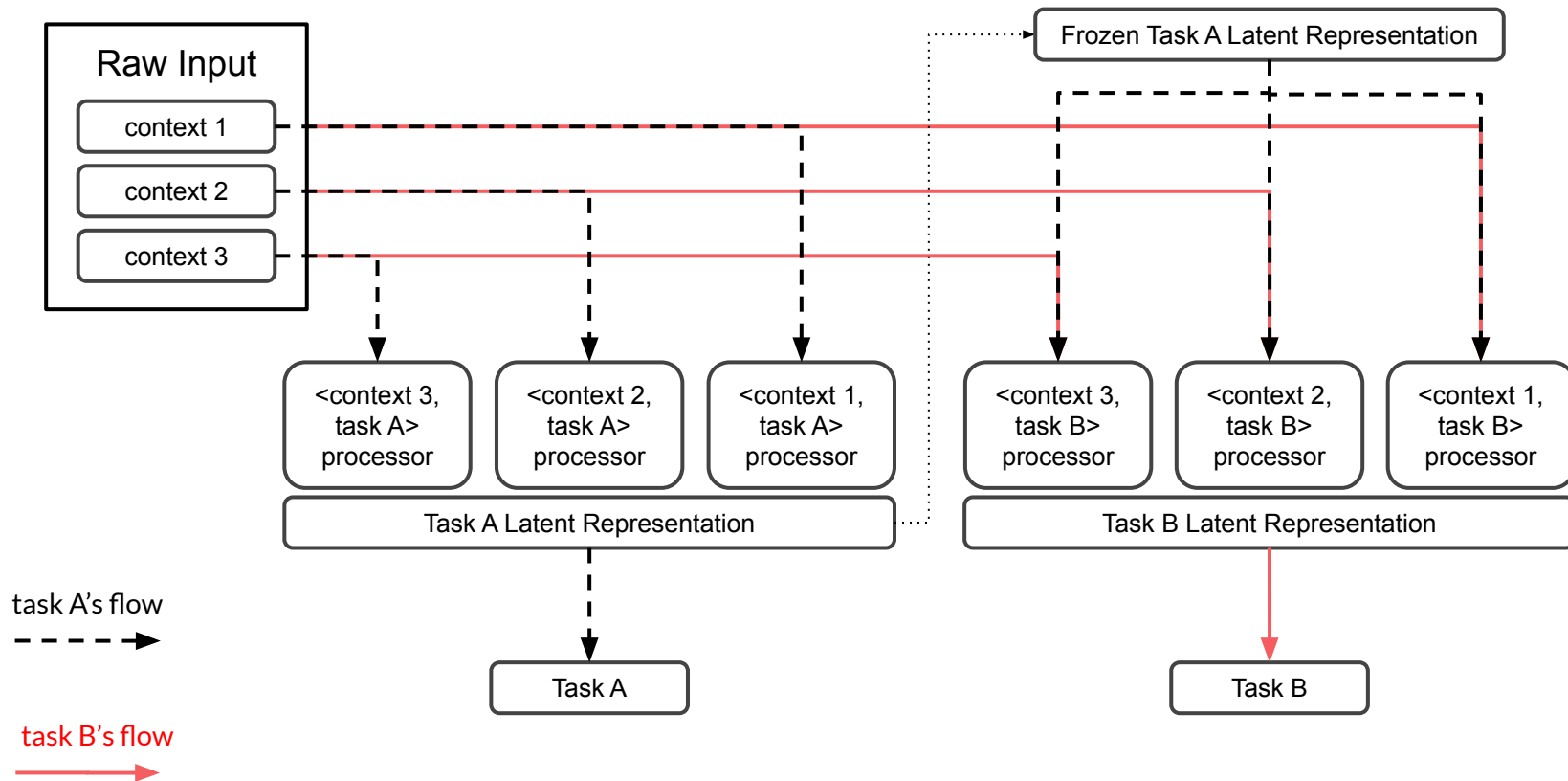


Task A is expressed in subjective human terms, e.g. telling cats and dogs apart.

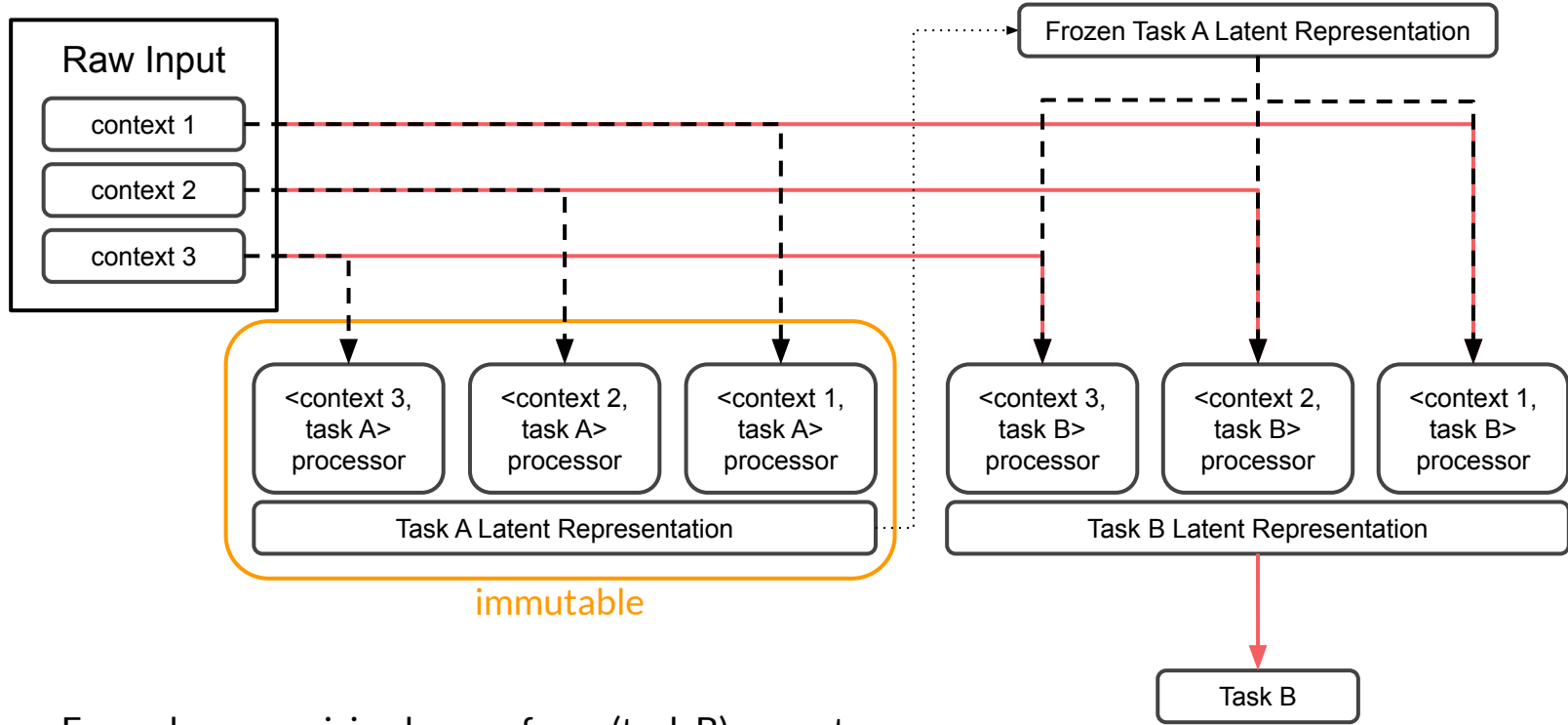
For $\text{HumanMeaning}(\text{TaskA}) = \text{Meaning}(\text{Task A Latent Representation})$ to hold true, task A's classes must be accurately predicted* across many contexts. If there aren't enough contexts, there is no guarantee that $\text{Meaning}(\text{Task A Latent Representation})$ aligns with $\text{HumanMeaning}(\text{TaskA})$.

*prediction of classes or numerical values, or rewards from motor goals.

Two-task scenario

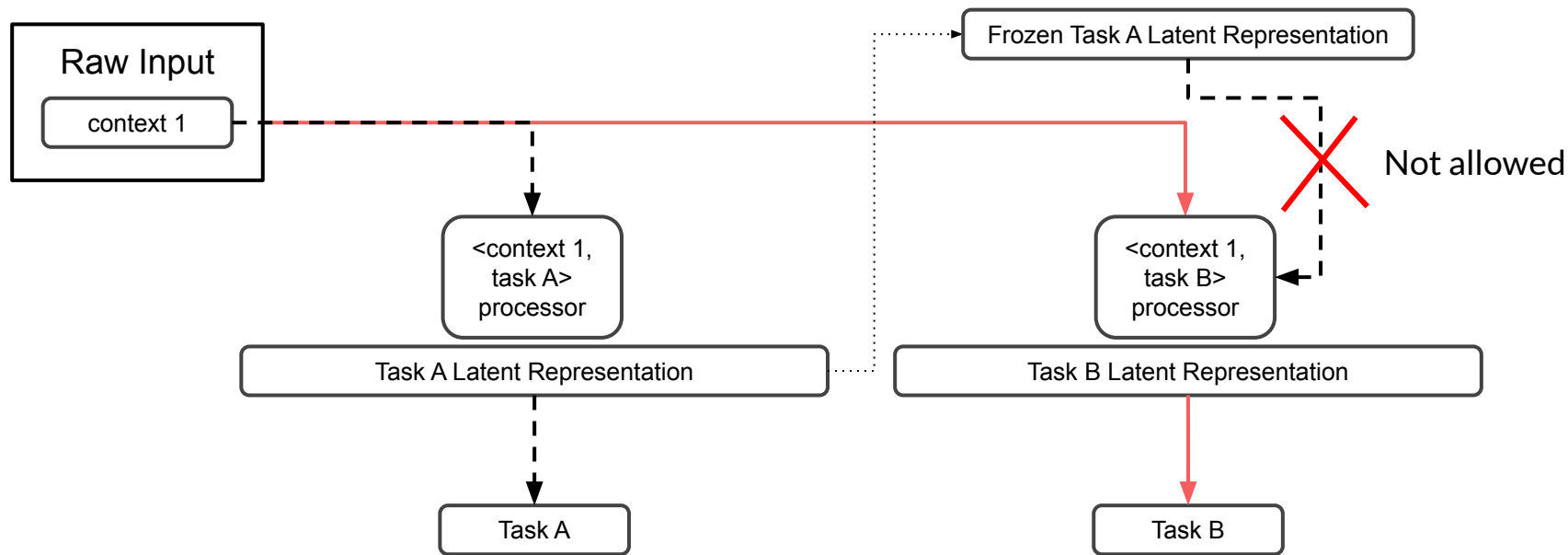


Rule 1: task A is frozen when training task B



Example: recognizing human faces (task B) cannot interfere with the task of telling cats and dogs apart (task A).

Rule 2: task B cannot utilize task A if training contexts were scarce

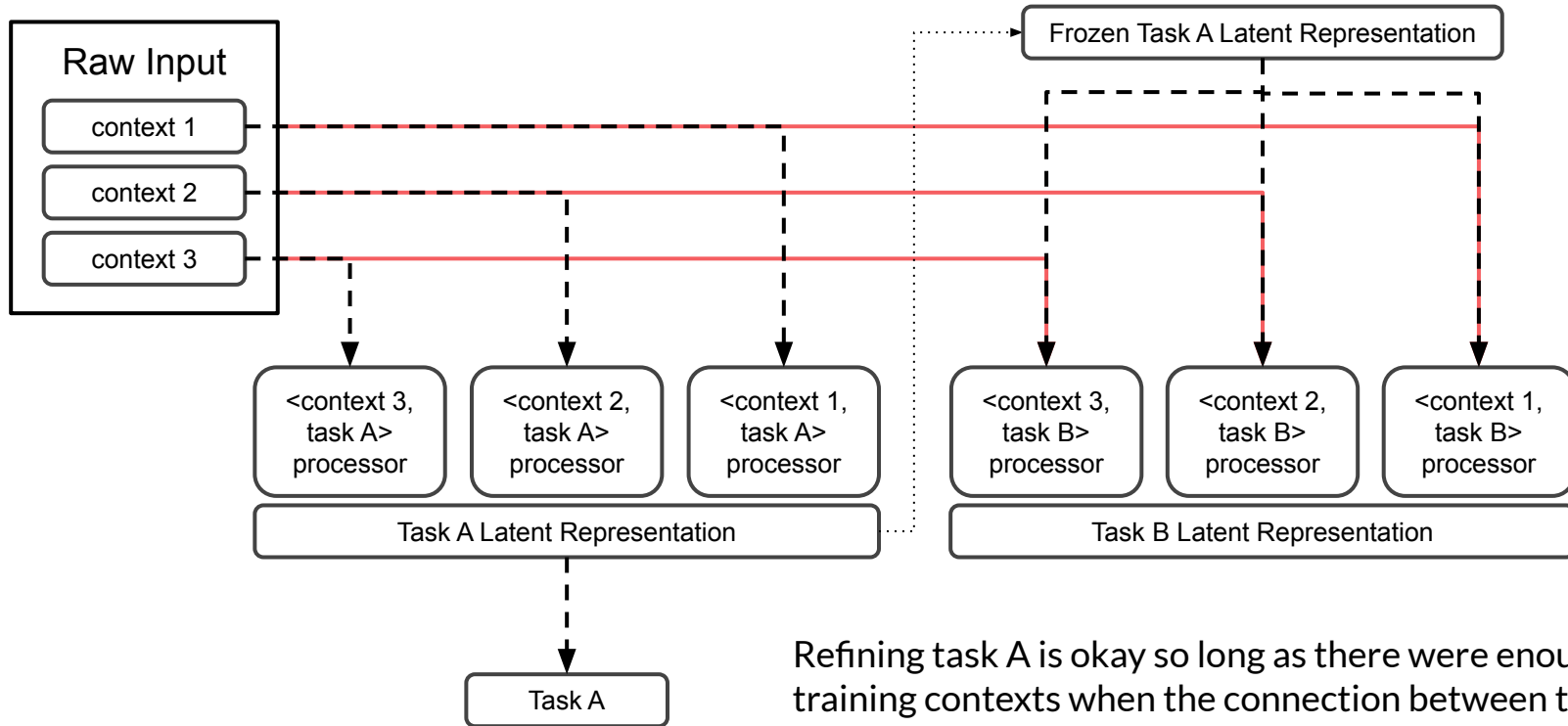


If there is a misalignment due to the lack of training contexts, i.e. $\text{HumanMeaning}(\text{TaskA}) \neq \text{Meaning}(\text{Task A LatentRepresentation})$,

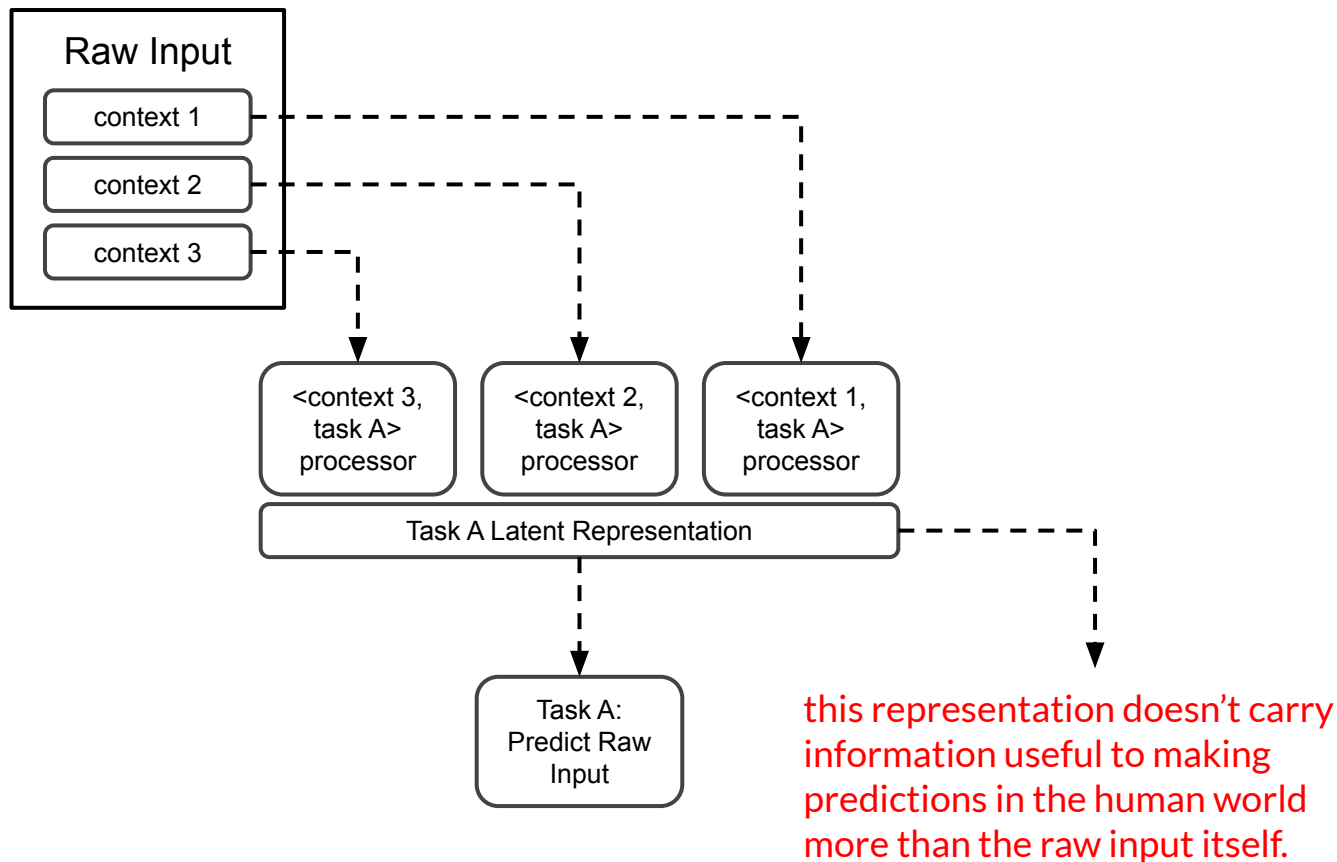
then the system might find a correlation between Task A and Task B that doesn't exist in human reality.

Example: without this rule, the system might mistakenly connect cats (task A) to arctic foxes (task B) if white cats were the only kind of cats seen by the system. If a connection between task A and B were to be drawn, nothing would stop task A from interfering with task B in a destructive way.

Rule 3: Task A is allowed to interfere on task B under rule 2's constraints

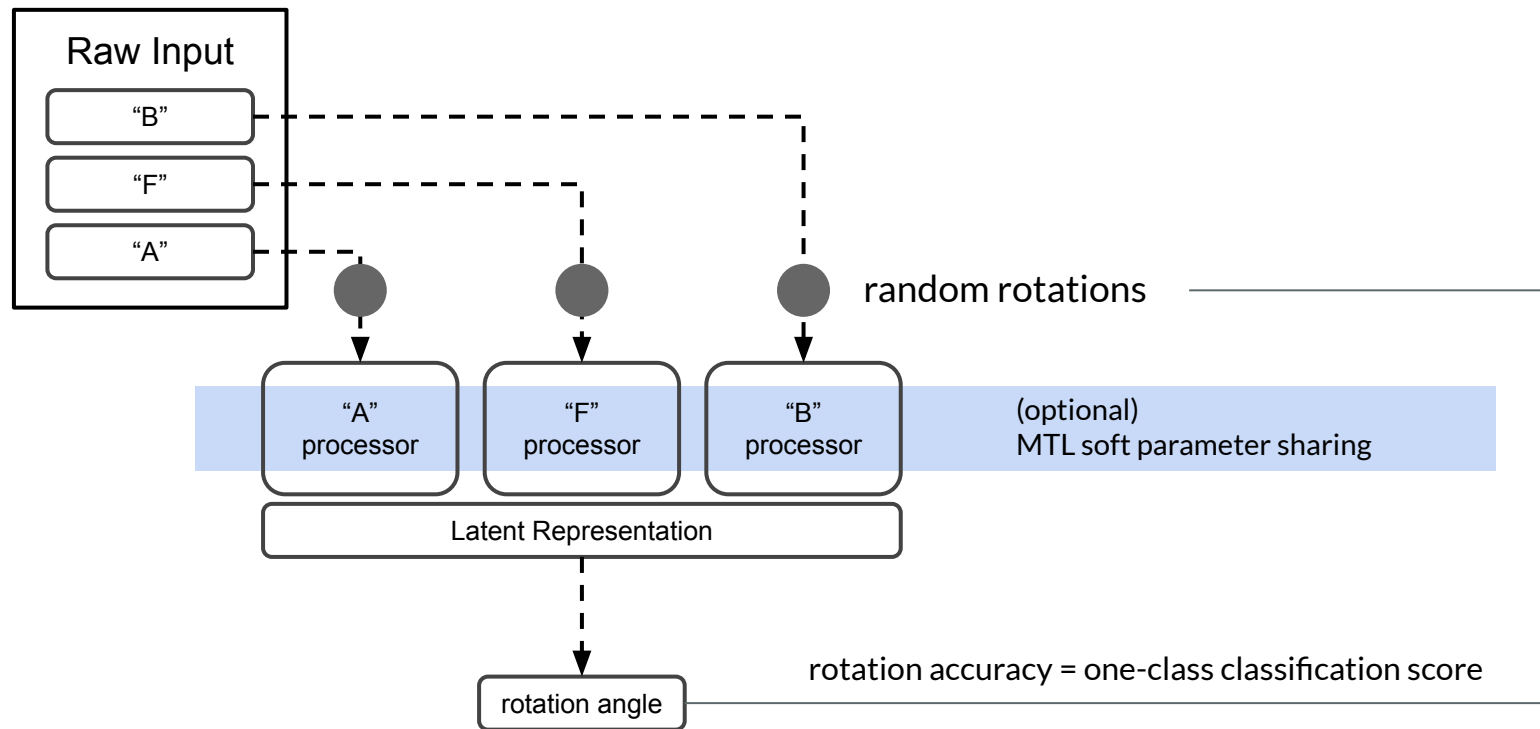


Rule 4: Tasks must interpret the raw input in subjective human terms



Autoencoding is an example of useless task.

One-task scenario: EMNIST



This works because "rotation" is a subjective concept that has the same meaning irrespective of the letter. It is hard to guess the rotation angle without customizing the process for each letter.

Example of invalid task: if the task is to denoise letter images, it can be done without knowing the letter, thus it doesn't help us predict the letter class.