Celem uczenie ze jest rozwiązywania problemów sekwencyjnych. W każdym kroku, agent obserwuje stan środowiska, w którym operuje i otrzymuje pewną nagrodę. Na podstawie tego decyduje o następnej akcji jaką będzie wykonywał. Środowisko przechodzi do następnego kroku, agent obserwuje nowy stan itp/itd. Celem jest znalezienie polityki, czyli sposobu generowania akcji, takiego, które wygeneruje maksymalną łączną nagrodę w epizodzie. Polityka jest funkcją ze stanów w akcje, w przypadku głębokiego uczenia ze wzmocnieniem jest realizowana przez głęboką sieć neuronową. Najbardziej znany przykładami problemów rozwiązywanych za pomocą uczenia ze wzmocnieniem są gry. Tym niemniej formalizm jest bardzo pojemy pozwalając opisać wiele problemów, np. pojazdy autonomiczne, optymizację sieci energetycznych itd. Niektóry sądzą, że uczenie ze wzmocnieniem może być drogą ku prawdziwie inteligentnym systemom.

Te nadziej spowodowały bardzo szybki rozwój dziedziny, przynosząc spektakularne sukcesy. Najbardziej znane to sukcesy w grach Dota2 i Starcraft. Sukcesy te budzą jednak pewne obiekcje. Metody uczenia ze wzmocnieniem wymagają bardzo dużej ilości doświadczenia i mocy obliczeniowej (np. w treningu Doty2 używano 2000 GPU i 100 tyś procesorów). Jest to problemem nie tylko z praktycznego punktu widzenia, ale też wskazuje na głębsze problemy. Wielu, włączając to luminarzy dziedziny, podnosi potrzebę rewizji podstawowych założeń dziedziny.

Podobne historie są powszechne w nauce. Po okresie bardzo szybkiego wzrostu jest potrzeba zastanowienia się nad fundamentami i tym samym ustalenia nowych kierunków badawczych. Ten projekt wpisuje się w ten prąd. Planuje wnieść kontrybucje w dwóch pod-dziedzinach: nienadzorowanym uczeniu ze wzmocnieniem i wielo-zadaniowym uczeniu ze wzmocnieniem. Obie z tych dziedzin mają za zadanie dostarczyć 'lepszego sygnału uczącego'. Intuicyjnie agent używający jedynie podstawowego sygnału nagród ma małą zachętę, aby poznać środowisko lepiej. Można sobie wyobrazić, że uczący się agent jest bardzo leniwym uczniem, który wykonuje minimalny wysiłek, aby rozwiązać zadanie. Takie zachowanie jest niepożądane, często prowadzi do suboptymalnych zachowań, które słabo się transferują nawet w przypadku podobnych środowisk.