Na przestrzeni kilku ostatnich lat sztuczna inteligencja (SI), a w szczególności uczenie głębokie, bez wątpienia zrewolucjonizowała branżę technologiczną. Wpływ SI jest również coraz częściej zauważalny w życiu codziennym. Mówiąc o sztucznej inteligencji, najczęściej mamy na myśli głębokie sieci neuronowe (ang. Deep Neural Networks, DNN), które potrafią wykonywać niezwykle złożone zadania, poprzez naśladowanie struktury ludzkiego mózgu. Klasyczny sposób trenowania sieci głębokiej polega na pokazywaniu jej milionów przykładów zawierających pytanie i odpowiedź, zwanych danymi i etykietami. Przekazując na wejście sieci pytania i porównując jej odpowiedzi z prawdziwymi wartościami, możemy uczyć algorytm tak, aby stopniowo stawał się coraz lepszy w rozwiązywaniu danego problemu. Takie podejście nazywa się uczeniem nadzorowanym i jest bardzo skuteczne w wielu pojedynczych zadaniach, dla których dostępny jest duży zbiór danych treningowych z odpowiednimi etykietami. W rzeczywistych zastosowaniach często jednak nie mamy dostępu do etykiet lub innych treści referencyjnych, co uniemożliwia korzystanie z metod uczenia nadzorowanego. Powodem braku etykiet może być unikalność problemu, bądź zbyt wysoki koszt etykietowania, na ogół wykonywanego ręcznie. Innym poważnym problemem jest brak tzw. danych wewnątrz-dziedzinowych. Problem ten występuje, gdy dane treningowe różnią się znacznie od danych napływających do systemu w trakcie jego działania. Przykładowo taka sytuacja ma miejsce, gdy sieć neuronowa jest trenowana za pomocą nagrań wykonanych profesjonalnym mikrofonem w warunkach studyjnych, a test odbywa się na podstawie nagrania z mikrofonu smartfona na ruchliwej ulicy.

Samouczenie (ang. self-supervision) stanowi rozwiązanie tych problemów. Polega bowiem na wymuszeniu, by sieć, w czasie treningu, samoistnie wykonywała powierzone zadanie (np. klasyfikacji) na podstawie struktur wewnątrz danych, w przeciwieństwie do treningu wykorzystującego bezpośrednio pary pytanie-odpowiedź, jak w uczeniu nadzorowanym. Dzięki temu możliwe jest wykorzystanie do nauki zbiorów bez etykiet, których koszt uzyskania jest znacznie mniejszy. Ponadto, samouczenie pozwala na dokładniejsze modelowanie danych, co ma pozytywny wpływ na działanie algorytmu w wielu późniejszych zastosowaniach. Zwykle osiąga się to poprzez włączenie do procesu uczenia etapu przewidywania. Na przykład, w przetwarzaniu języka naturalnego (ang. Natural Language Processing) sieć w trakcie treningu otrzymuje zdania pozbawione niektórych słów (słowa te zostają zakryte), które są następnie przewidywane wyłącznie na podstawie kontekstu.

Do tej pory podjęto jedynie ograniczone próby wprowadzenia samouczenia do zastosowań akustycznych, dlatego celem projektu Acoustic Intelligence jest zbadanie jego pełnego potencjału w tym kontekście. Planowane jest dokładne zbadanie różnych aspektów samouczenia, z czego jednym z proponowanych nowatorskich podejść jest trening sieci umożliwiającej ekstrakcję uniwersalnej reprezentacji dźwięku, którą można wykorzystać w wielu niezależnych zadaniach. Zakłada się, że dzięki wykorzystaniu samouczenia w trakcie treningu sieci do wykonywania wielu zadań jednocześnie, jesteśmy w stanie poprawić jej zdolność do uchwycenia wiedzy ogólnej i bazującej na wewnętrznej strukturze występującej w sygnałach akustycznych. Innym przypadkiem, w którym spodziewamy się zysku spowodowanego wykorzystaniem samouczenia jest analiza sceny dźwiękowej, w której detekcja konkretnych zdarzeń akustycznych będzie wpływała na klasyfikację sceny akustycznej, np. śpiew ptaków i bawiące się dzieci mogą pomóc zakwalifikować scenę jako park. Planowane jest użycie tych semantycznych związków między zdarzeniami akustycznymi poprzez zawarcie technik przetwarzania języka naturalnego. Pozostała część projektu skupiać się będzie na uniwersalnej poprawie jakości sygnału dźwiękowego oraz separacji dźwięków. Wprowadzając nowe techniki, takie jak samouczenie czy uczenie poprzez wzmocnienie (ang. reinforcement learning), jak również wykorzystanie informacji z wielu mikrofonów, chcemy zminimalizować wymóg stosowania sygnałów referencyjnych w treningu i w szerokim kontekście podnieść wydajność technik poprawy jakości sygnałów akustycznych opartych na uczeniu głębokim. Oczekujemy, że w wyniku realizacji tego projektu w znaczący sposób przyczynimy się do rozwoju badań nad samouczeniem w zastosowaniach akustycznych, osiągając wyniki wykraczające poza granice tego co jest możliwe dziś przy użyciu standardowego uczenia nadzorowanego.

Wyniki projektu Acoustic Intelligence mogą mieć realny wpływ na życie codzienne. Zaproponowane rozwiązania mogą posłużyć do usprawnienia systemów wspomagania słuchu takich jak inteligentne aparaty słuchowe, mogą być też wykorzystane do ostrzegania przed potencjalnie niebezpiecznymi sytuacjami na drodze, jak choćby o zbliżającej się karetce pogotowia. Dzięki nowym modelom potężne zbiory audio będą mogły być automatycznie opisane słownie oraz przeszukiwane pod kątem konkretnej zawartości, na przykład sceny lub osoby.