Głębokie uczenie się osiągnęło znaczące sukcesy w wielu dziedzinach, takich jak generowanie obrazów, przetwarzanie języka naturalnego, zwijanie białek czy projektowanie materiałów. Pomimo różnic w architekturach i strategiach optymalizacji, te przełomy opierają się na fundamentalnym założeniu: stałym zbiorze danych treningowych. Obecnie, złamanie tego założenia prowadzi do wielu problemów, takich jak katastrofalne zapominanie czy utrata plastyczności. Jednakże, wykroczenie poza paradygmat stałego zbioru danych umożliwiłoby ciągłą adaptację tych modeli do stale zmieniającego się środowiska — cechy charakterystycznej dla każdej inteligentnej jednostki. Aby to umożliwić, musimy odpowiedzieć na kilka fundamentalnych pytań badawczych. Pięć publikacji zawartych w tej pracy doktorskiej częściowo odpowiada na następujące pytanie:

Jaką rolę odgrywają reprezentacje danych w optymalizacji niestacjonarnej?

Praca [1] pokazuje, że modyfikacja reprezentacji danych przed treningiem może zmniejszyć katastrofalne zapominanie. W tym celu model-nauczyciel meta-uczy się syntetycznych próbek danych zaprojektowanych tak, aby złagodzić negatywne skutki niestacjonarnego treningu modeli-uczniów. Praca podkreśla znaczenie reprezentacji danych w kontekście katastrofalnego zapominania. Jednakże, koszty meta-uczenia sprawiają, że to podejście jest niepraktyczne. Potencjalne rozwiązanie tego problemu przedstawiono w pracy [2], która modyfikuje reprezentacje danych poprzez uczenie się niezmienników danych przy użyciu zastępczej funkcji straty rekonstrukcji, co wiąże się z pomijalnym dodatkowym kosztem. Eksperymenty pokazują, że uczenie się tych niezmienników zwiększa odporność modelu na katastrofalne zapominanie.

Ograniczenia prac [1, 2] oraz wielu innych prac z zakresu ciągłego uczenia się wynikają z traktowania sieci neuronowych jako czarnych skrzynek, które przyjmują dane i zwracają przewidywania. Aby poszerzyć tę wąską perspektywę, prace [3, 4] koncentrują się na ukrytych reprezentacjach sieci neuronowych i badają ich wpływ na niestacjonarny trening. Praca [3] wprowadza Hipotezę Efektu Tunelu, która mówi, że warstwy wystarczająco głębokich sieci dzielą się na dwie odrębne części, które w różny sposób przyczyniają się do wykonania zadania. Początkowe warstwy generują reprezentacje liniowo separowalne, podczas gdy kolejne warstwy, zwane tunelem, kompresują te reprezentacje. Tunele odgrywają istotną rolę w generalizacji modelu i ciągłym uczeniu się, co podważa dotychczasowe poglądy na temat katastrofalnego zapominania. Kontynuując ten wątek, kolejna publikacja [4] teoretycznie analizuje dynamikę reprezentacji, aby zrozumieć źródła Efektu Tunelu, oraz dowodzi zaskakującej zdolności funkcji softmax do zwiększania rangi reprezentacji, co prowadzi do wielu konsekwencji, takich jak kompresja czy gorsza wydajność na danych spoza dystrybucji.

Ostatnia praca [5] podważa obecne rozumienie optymalizacji niestacjonarnej i empirycznie demonstruje, że ciągły trening na zadaniach z tym samym rozkładem reprezentacji danych może prowadzić do katastrofalnego zapominania i utraty plastyczności.