

Direction-aware and scaled Wasserstein distance vs. Epochs (GPT-2)

