Grid Search: Hidden Size vs Depth vs Repr Dim (600k training steps)