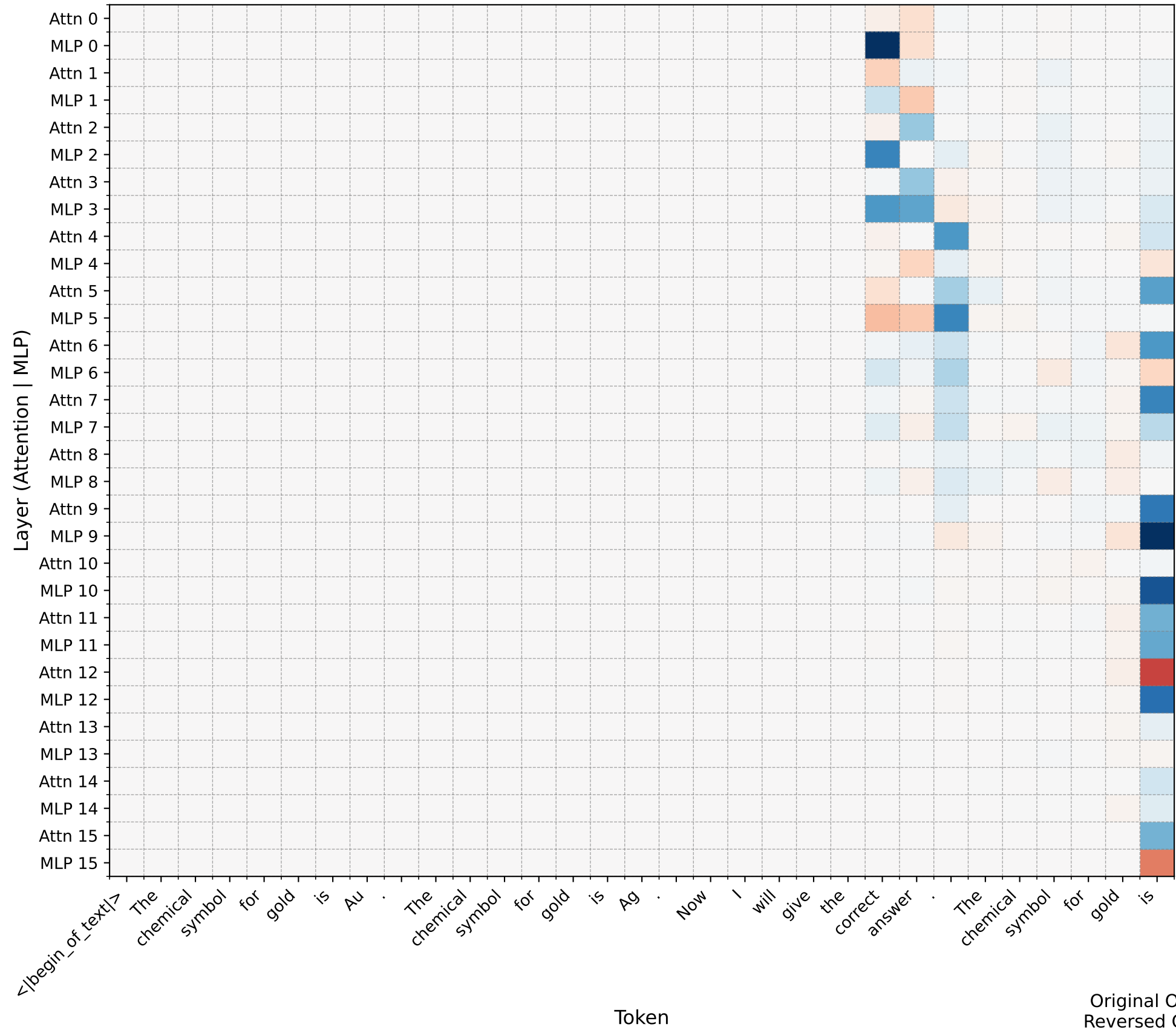


Patching Effects by Layer and Token
(Clean: "Paris" vs Corrupted: "Berlin")

Patching Effects (Original Order)



Patching Effects (Reversed Order)

