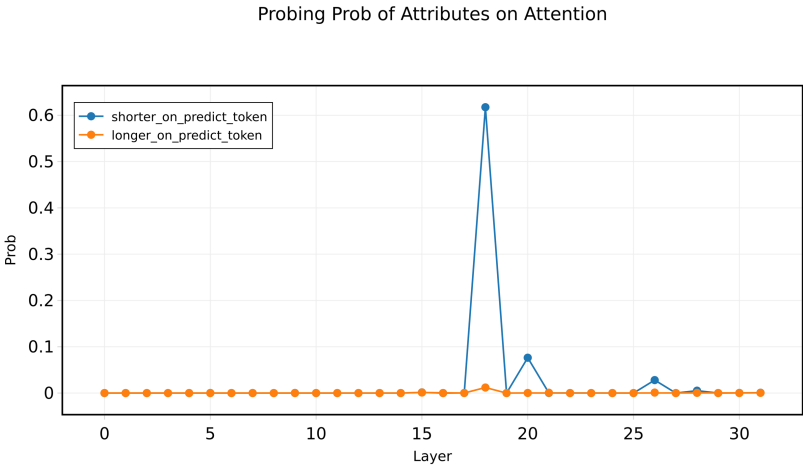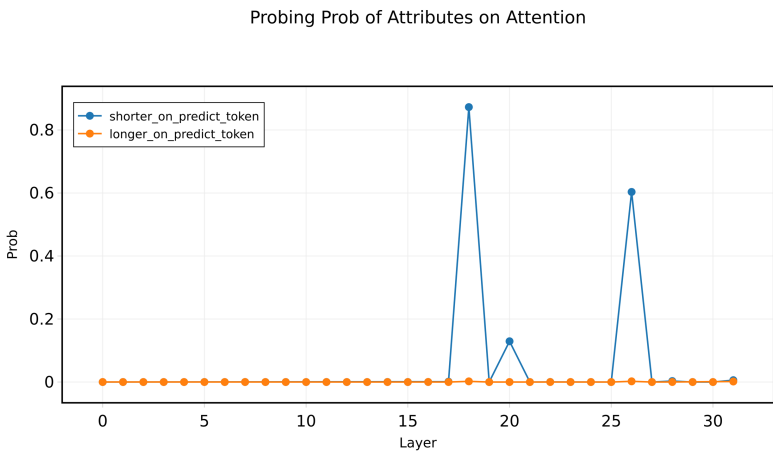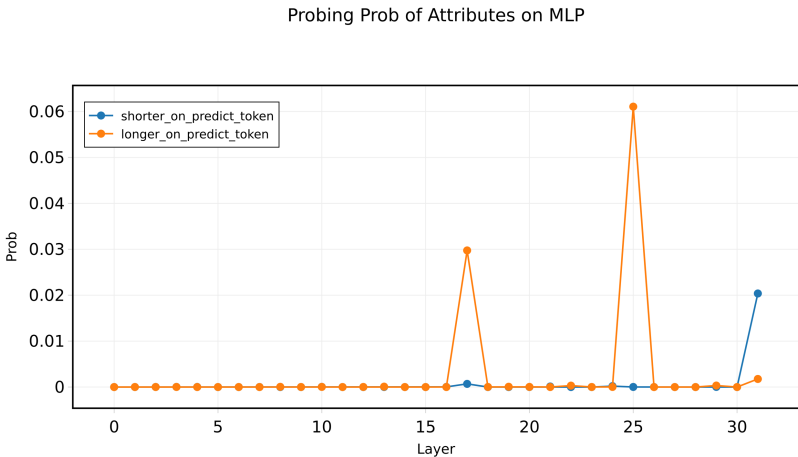| Input	Question: John cannot run the entire length of the track, he had been used to the field. The _ is short. (1) track (2) field Answer: A person who cannot run the entire length of a track likely feels uncomfortable or out of practice on a surface that is different from what they are used to. If John had been used to the field, it suggests that he is more accustomed to that environment. Therefore, the track must be
SSFT model output	**longer** than the field, making it difficult for him to run its entire length. ✅
Base model output	**shorter** than the field, as he struggles to run its entire length. ❌

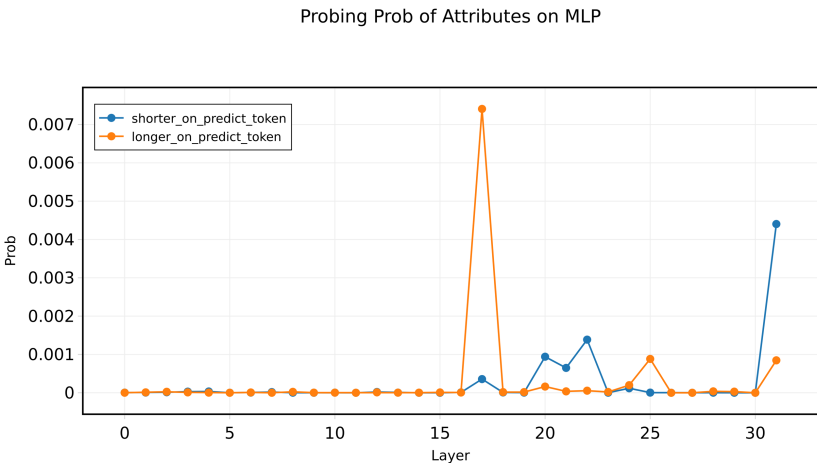(a) Case study: output of SSFT and Base model

(b) Probing attention layer output for "shorter" and "longer" on SSFT model

(c) Probing attention layer output for "shorter" and "longer" on Base model

(d) Probing MLP layer output for "shorter" and "longer" on SSFT model

(e) Probing MLP layer output for "shorter" and "longer" on Base model