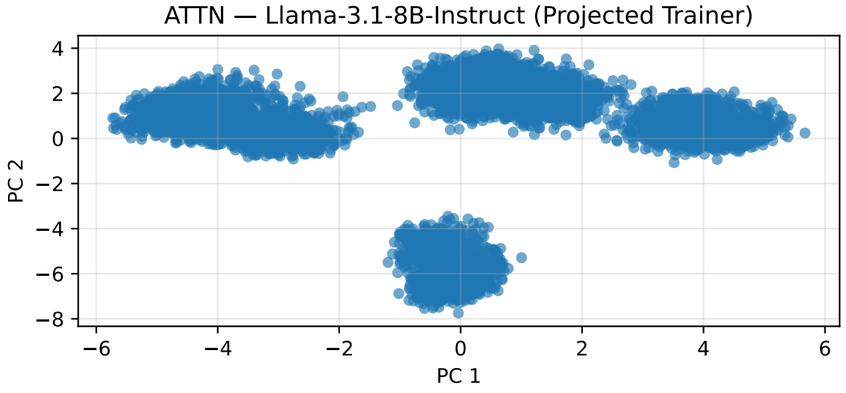
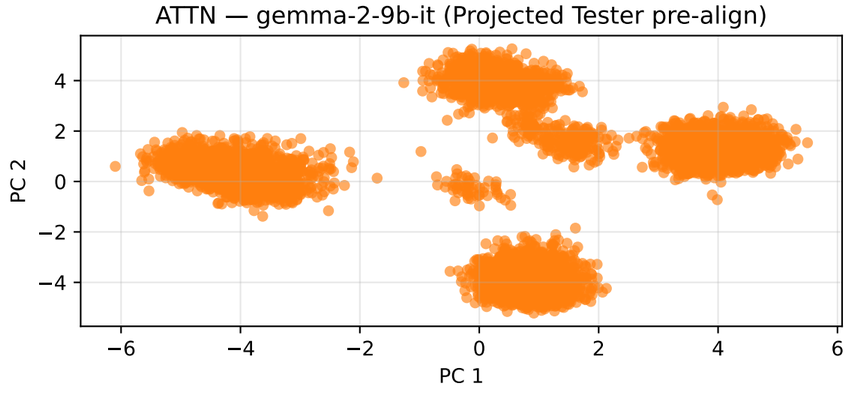
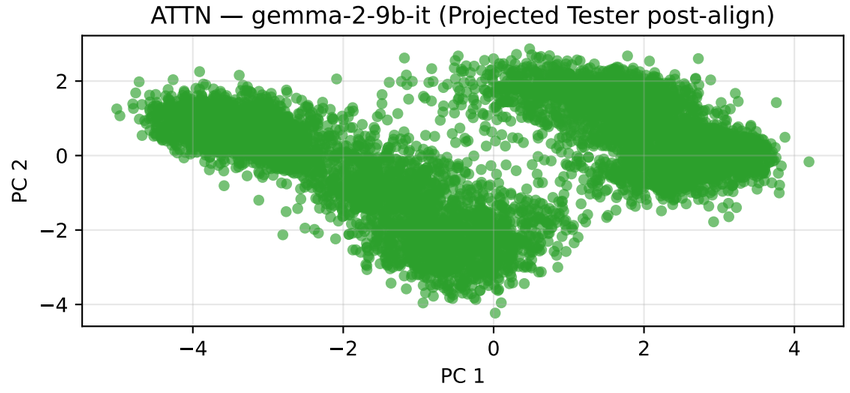
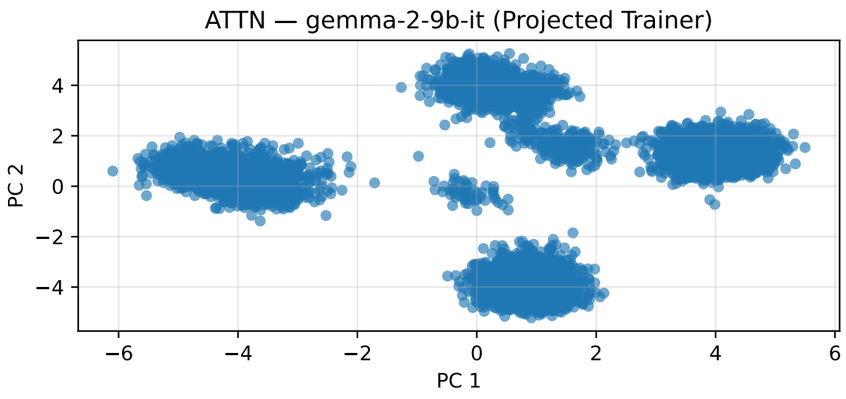
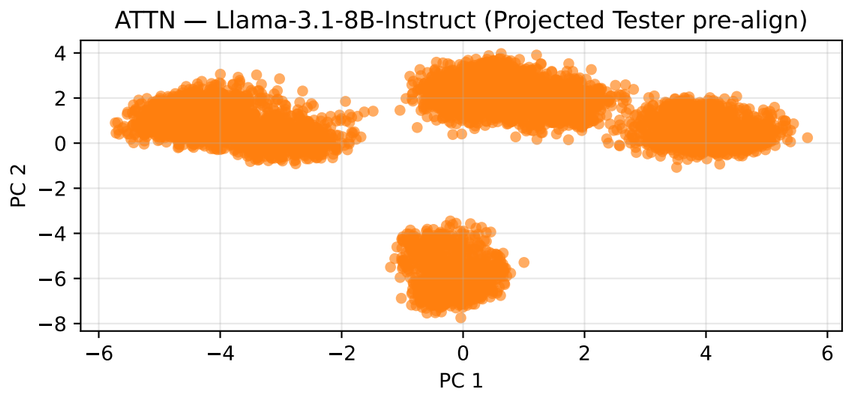
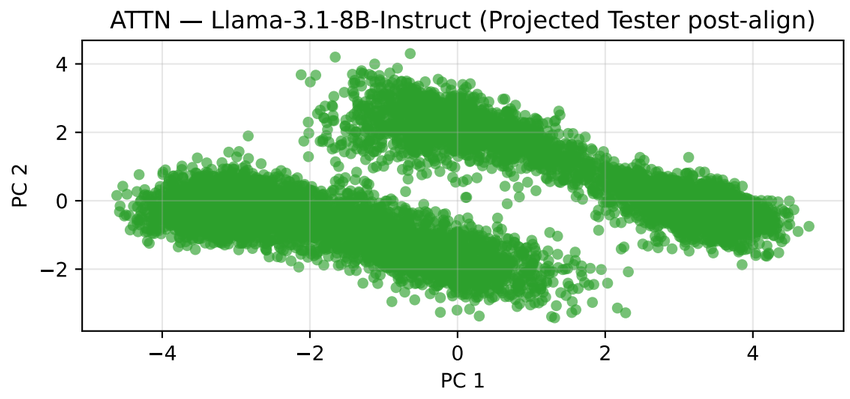

ATTN — Llama-3.1-8B-Instruct (Projected Trainer)
PC 1
PC 2

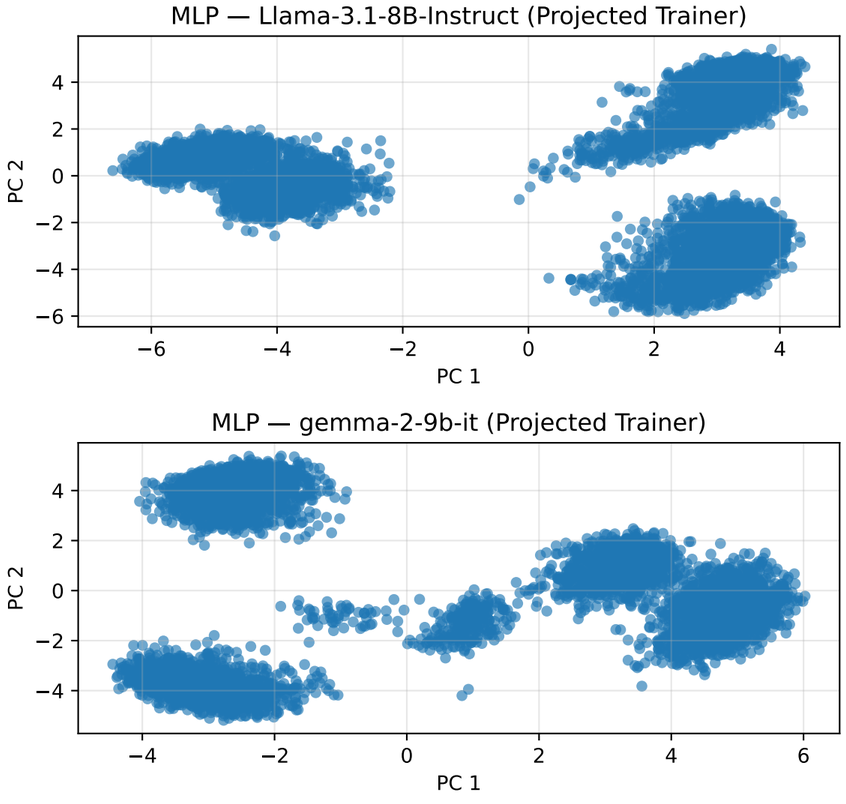

ATTN — gemma-2-9b-it (Projected Tester pre-align)
PC 1
PC 2

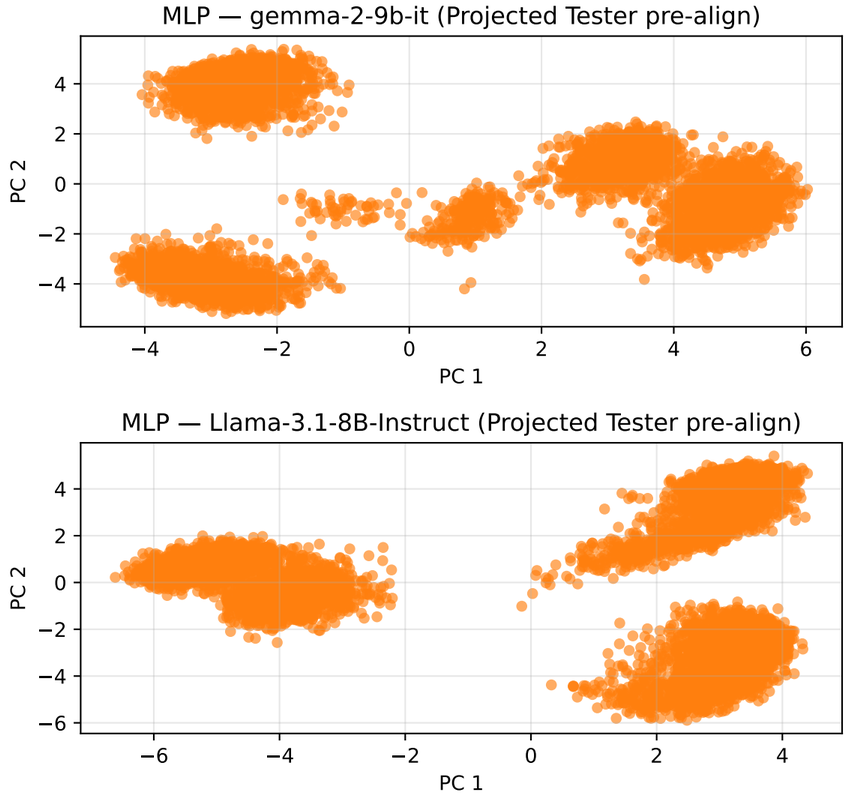

ATTN — gemma-2-9b-it (Projected Tester post-align)
PC 1
PC 2

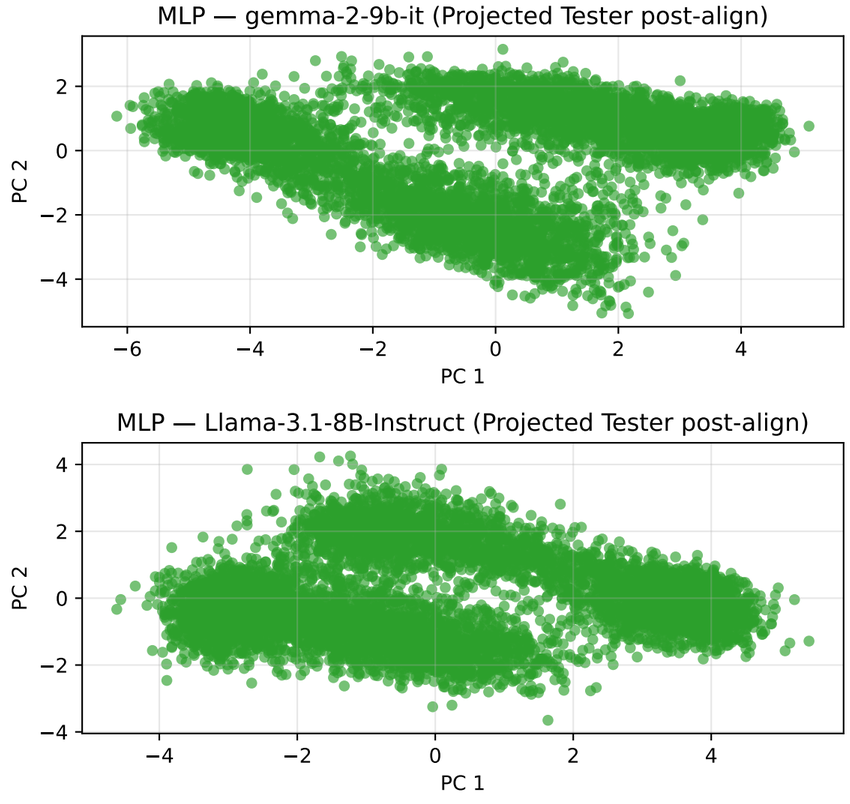

ATTN — gemma-2-9b-it (Projected Trainer)
PC 1
PC 2


ATTN — Llama-3.1-8B-Instruct (Projected Tester pre-align)
PC 1
PC 2


ATTN — Llama-3.1-8B-Instruct (Projected Tester post-align)
PC 1
PC 2


MLP — Llama-3.1-8B-Instruct (Projected Trainer)
PC 1
PC 2

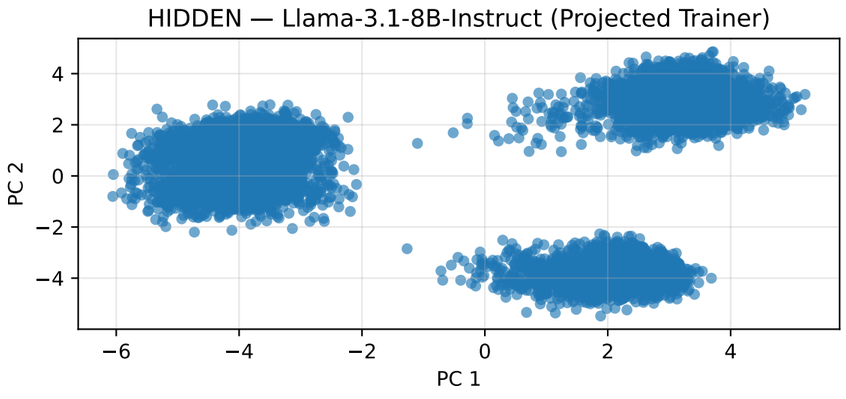

MLP — gemma-2-9b-it (Projected Tester pre-align)
PC 1
PC 2

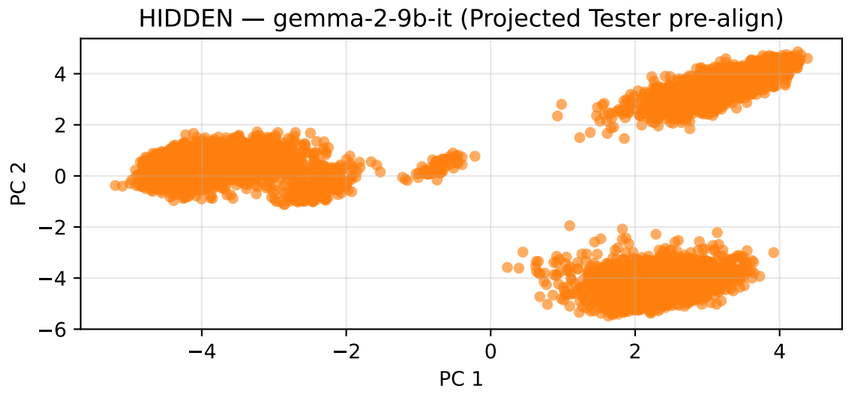

MLP — gemma-2-9b-it (Projected Tester post-align)
PC 1
PC 2

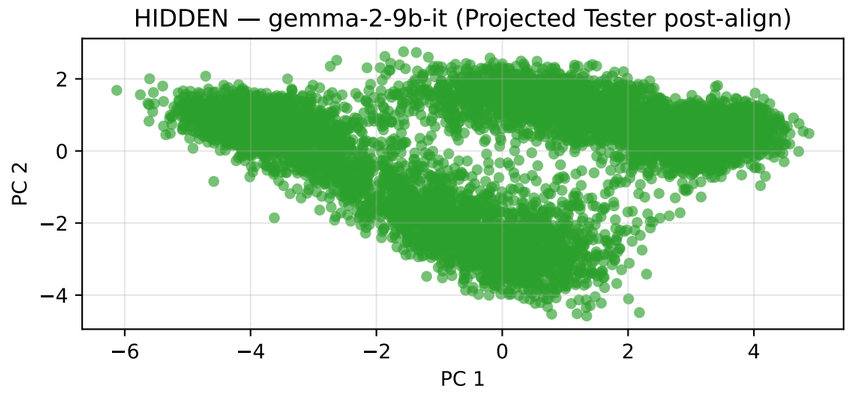
MLP — gemma-2-9b-it (Projected Trainer)
PC 1
PC 2

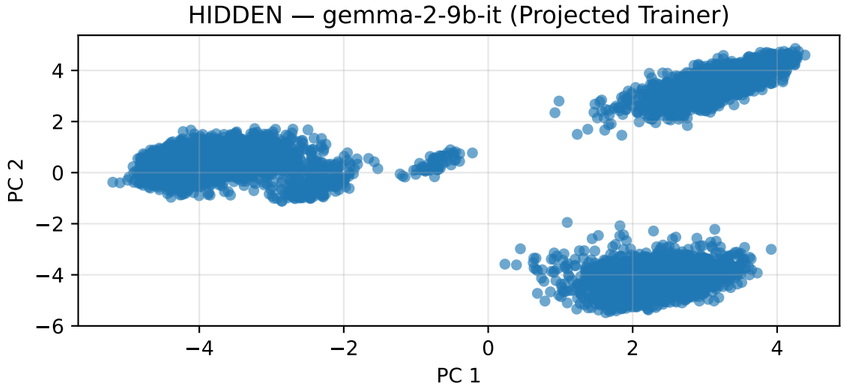
MLP — Llama-3.1-8B-Instruct (Projected Tester pre-align)
PC 1
PC 2

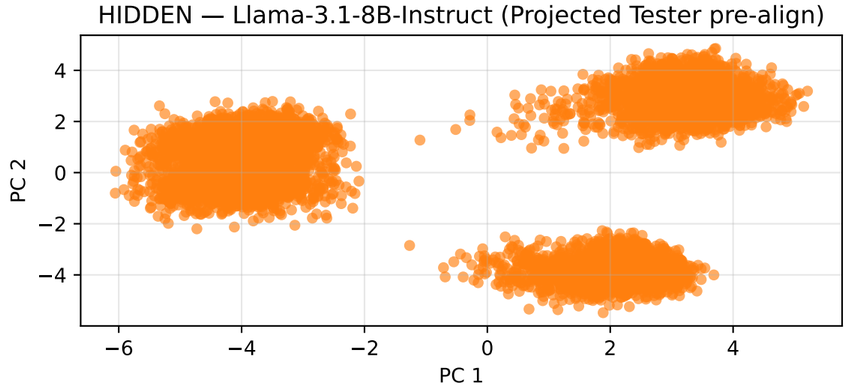
MLP — Llama-3.1-8B-Instruct (Projected Tester post-align)
PC 1
PC 2

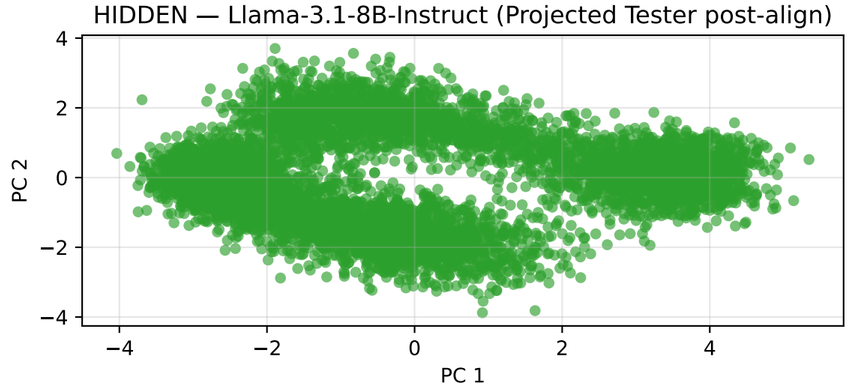

HIDDEN — Llama-3.1-8B-Instruct (Projected Trainer)
PC 1
PC 2


HIDDEN — gemma-2-9b-it (Projected Tester pre-align)
PC 1
PC 2


HIDDEN — gemma-2-9b-it (Projected Tester post-align)
PC 1
PC 2


HIDDEN — gemma-2-9b-it (Projected Trainer)
PC 1
PC 2


HIDDEN — Llama-3.1-8B-Instruct (Projected Tester pre-align)
PC 1
PC 2


HIDDEN — Llama-3.1-8B-Instruct (Projected Tester post-align)
PC 1
PC 2