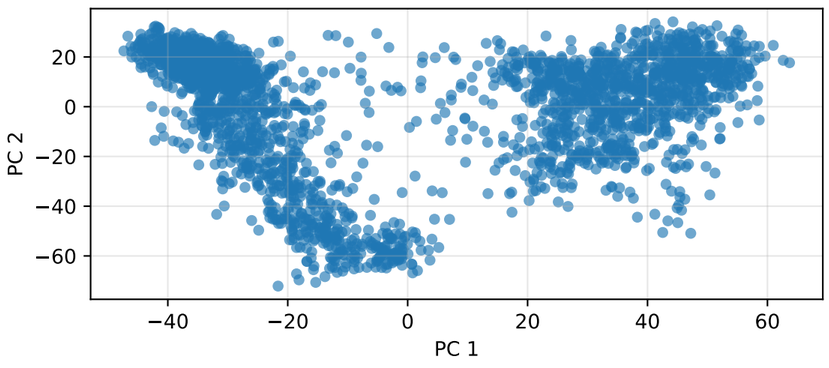
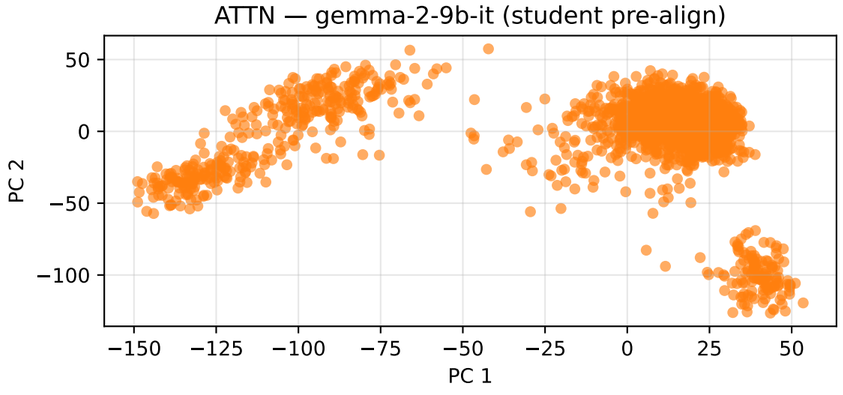
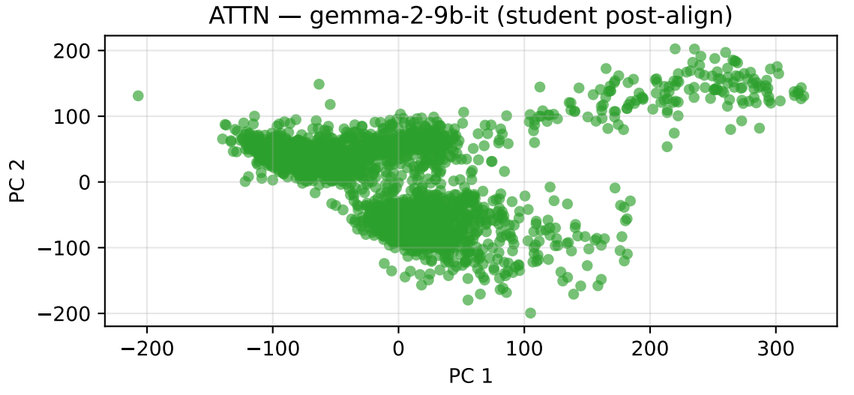
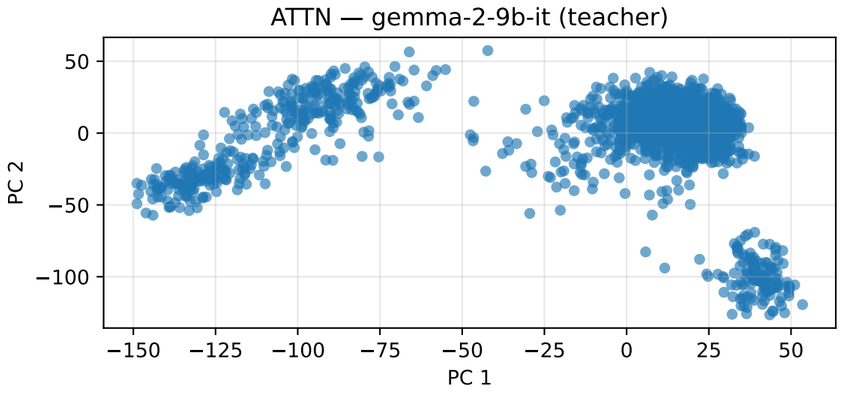
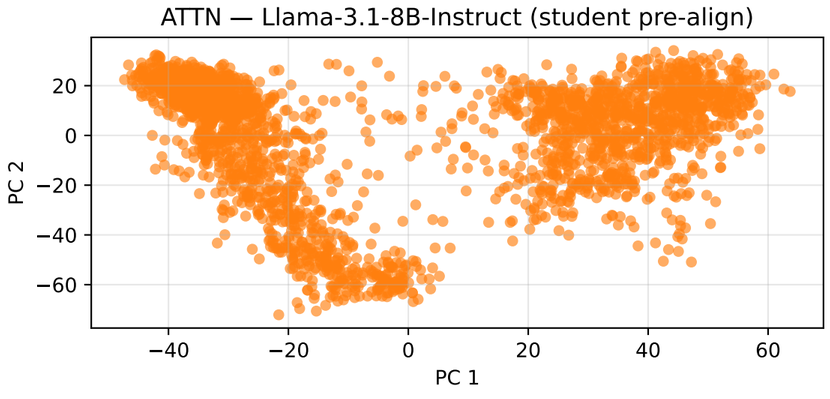
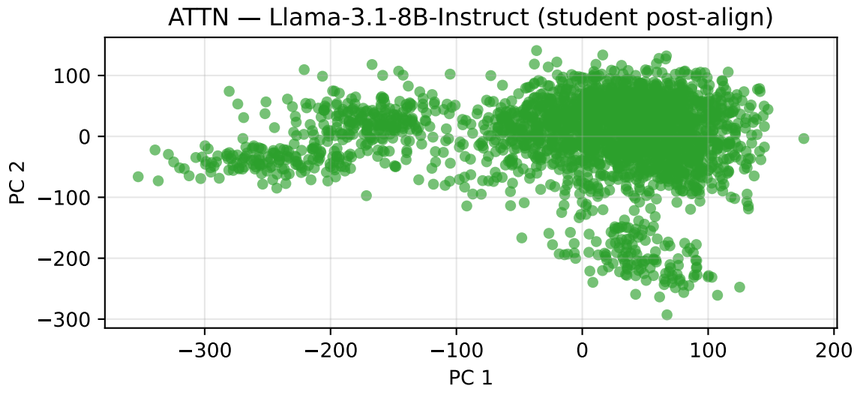

ATTN — Llama-3.1-8B-Instruct (teacher)
PC 1
PC 2

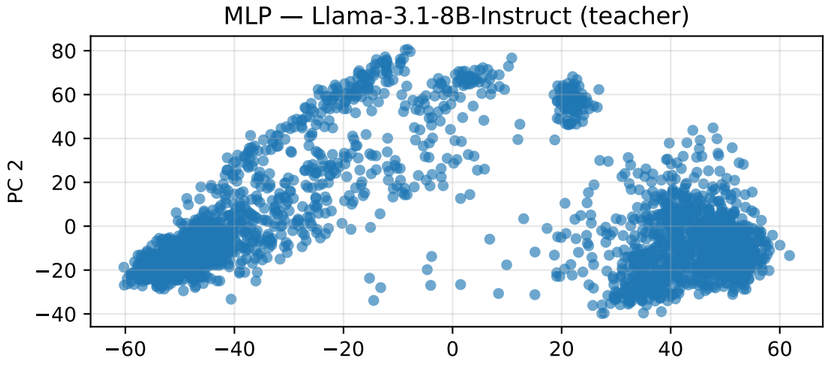

ATTN — gemma-2-9b-it (student pre-align)
PC 1
PC 2

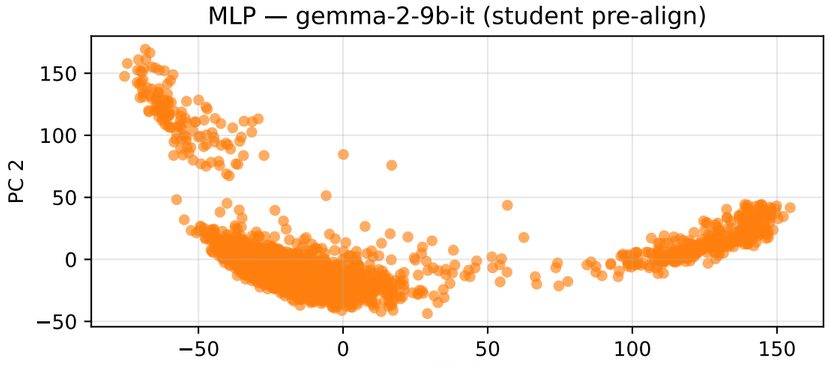

ATTN — gemma-2-9b-it (student post-align)
PC 1
PC 2

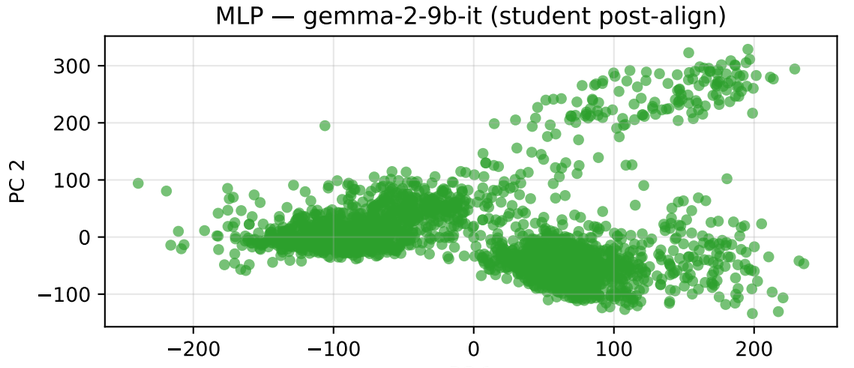

ATTN — gemma-2-9b-it (teacher)
PC 1
PC 2

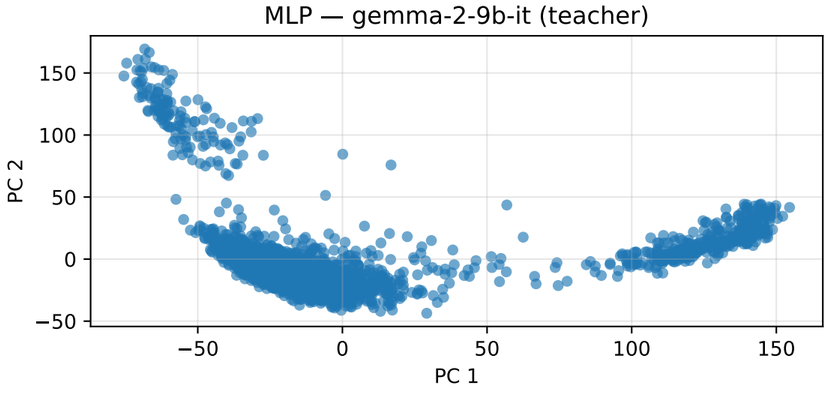

ATTN — Llama-3.1-8B-Instruct (student pre-align)
PC 1
PC 2

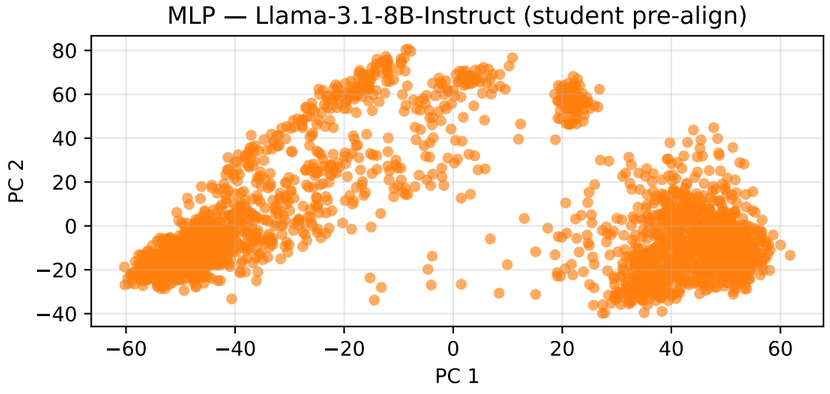

ATTN — Llama-3.1-8B-Instruct (student post-align)
PC 1
PC 2

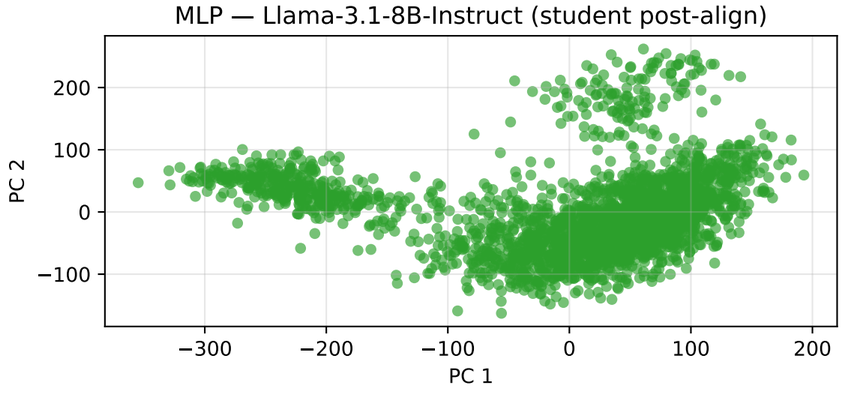

MLP — Llama-3.1-8B-Instruct (teacher)
PC 1
PC 2

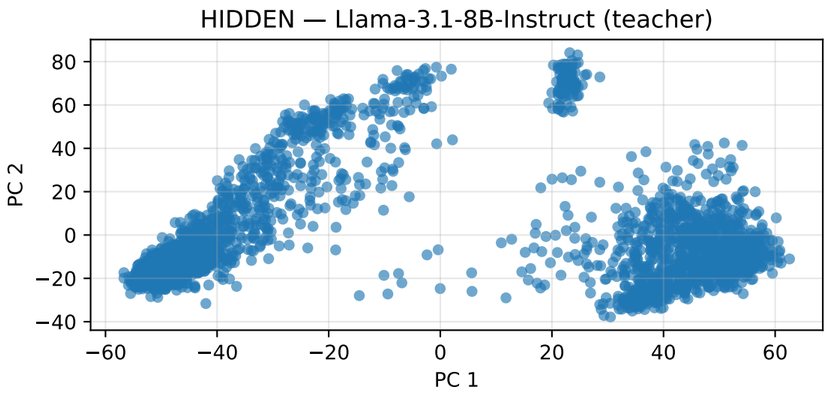

MLP — gemma-2-9b-it (student pre-align)
PC 1
PC 2

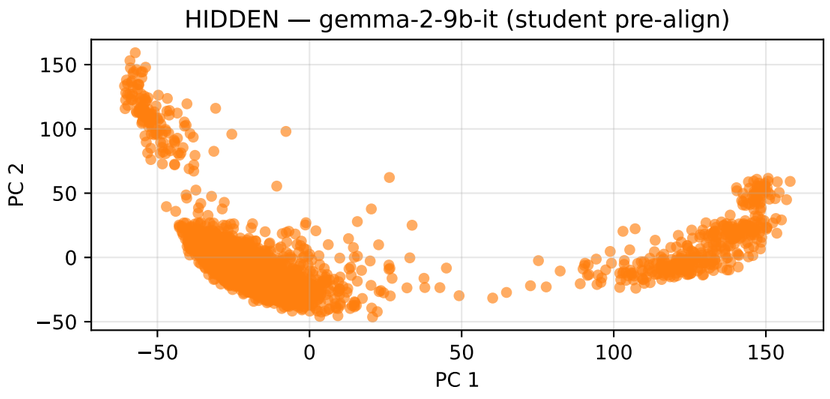

MLP — gemma-2-9b-it (student post-align)
PC 1
PC 2

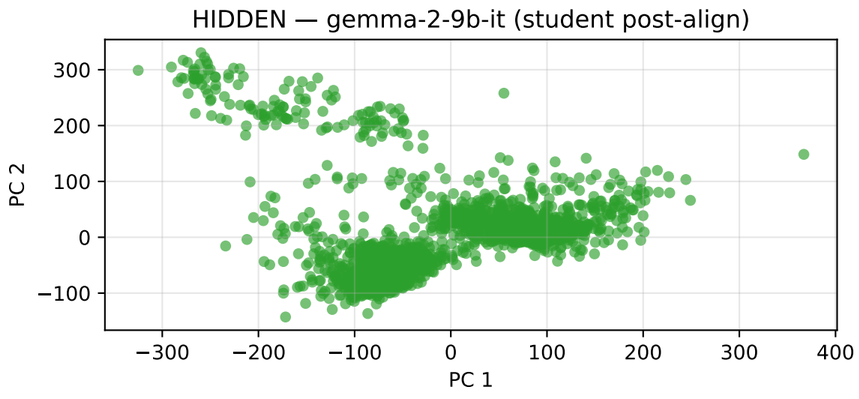

MLP — gemma-2-9b-it (teacher)
PC 1
PC 2

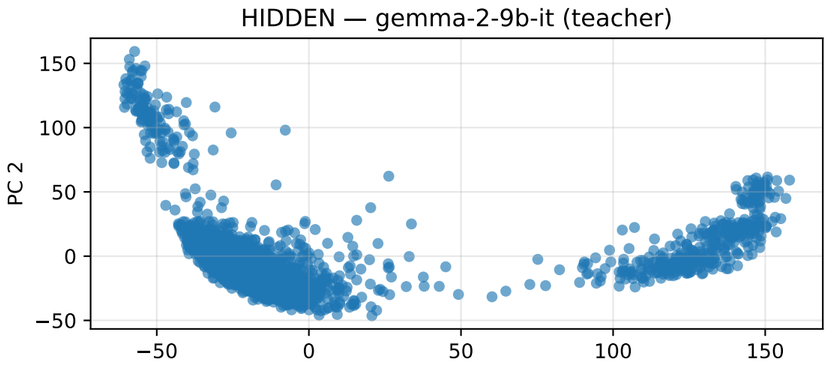

MLP — Llama-3.1-8B-Instruct (student pre-align)
PC 1
PC 2

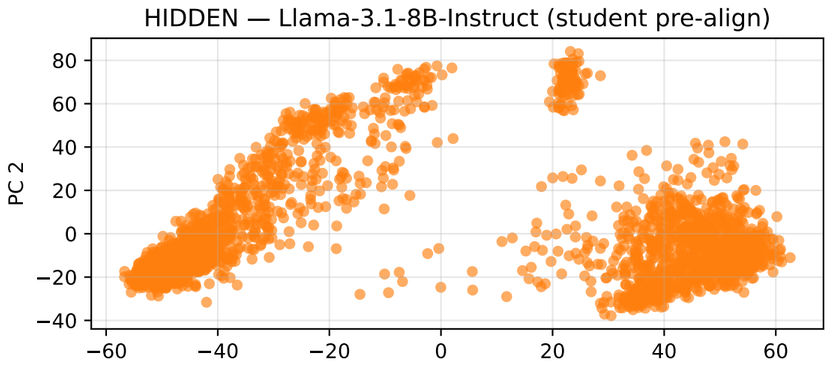

MLP — Llama-3.1-8B-Instruct (student post-align)
PC 1
PC 2

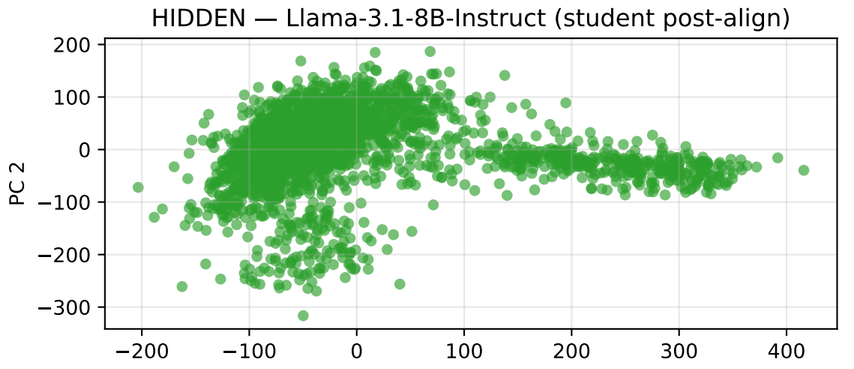

HIDDEN — Llama-3.1-8B-Instruct (teacher)
PC 1
PC 2


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


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


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


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


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