

Text Description

Diffusion Image

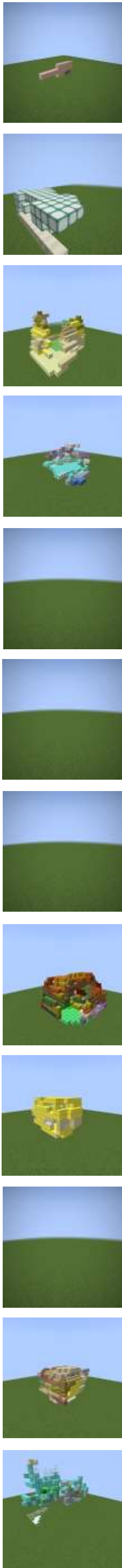
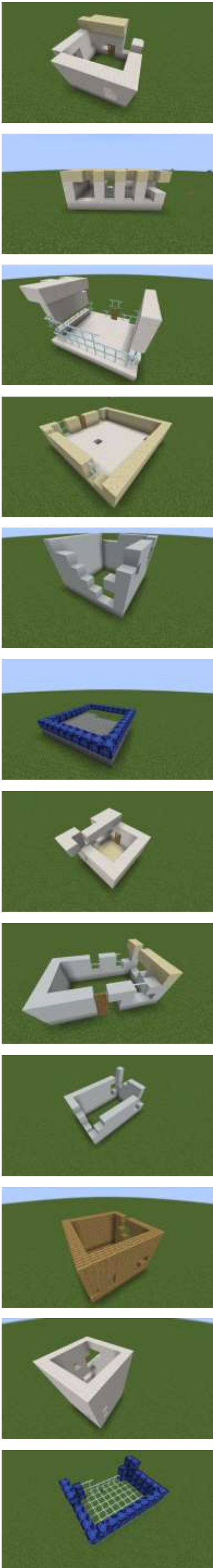
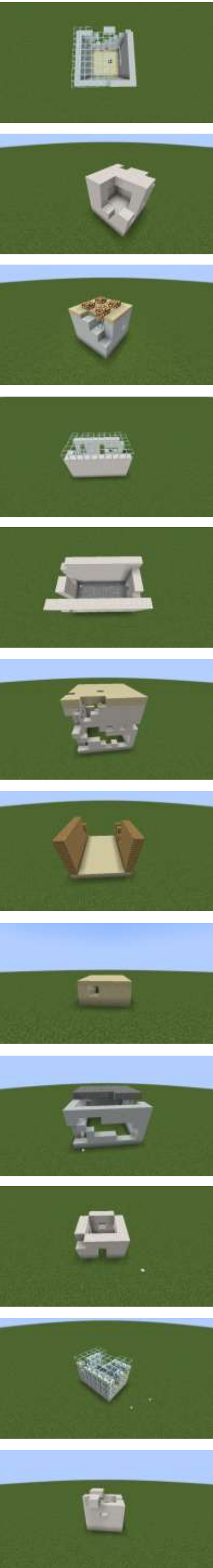
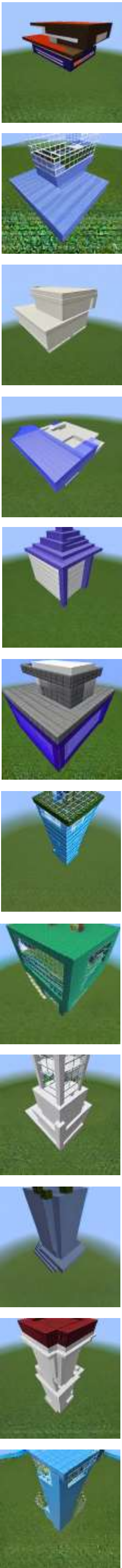
Vanilla GPT-4

CoT+GPT-4

Diffusion+GPT-4V

Diffusion+BC

“A modern building in Minecraft.”



“A tower in Minecraft.”