

**Input:** audio and single portrait image

Output: talking head animation