







- Sticthing Decoder Blocks together! - Named as GemmaModel

- Takes care of weight-tying technique