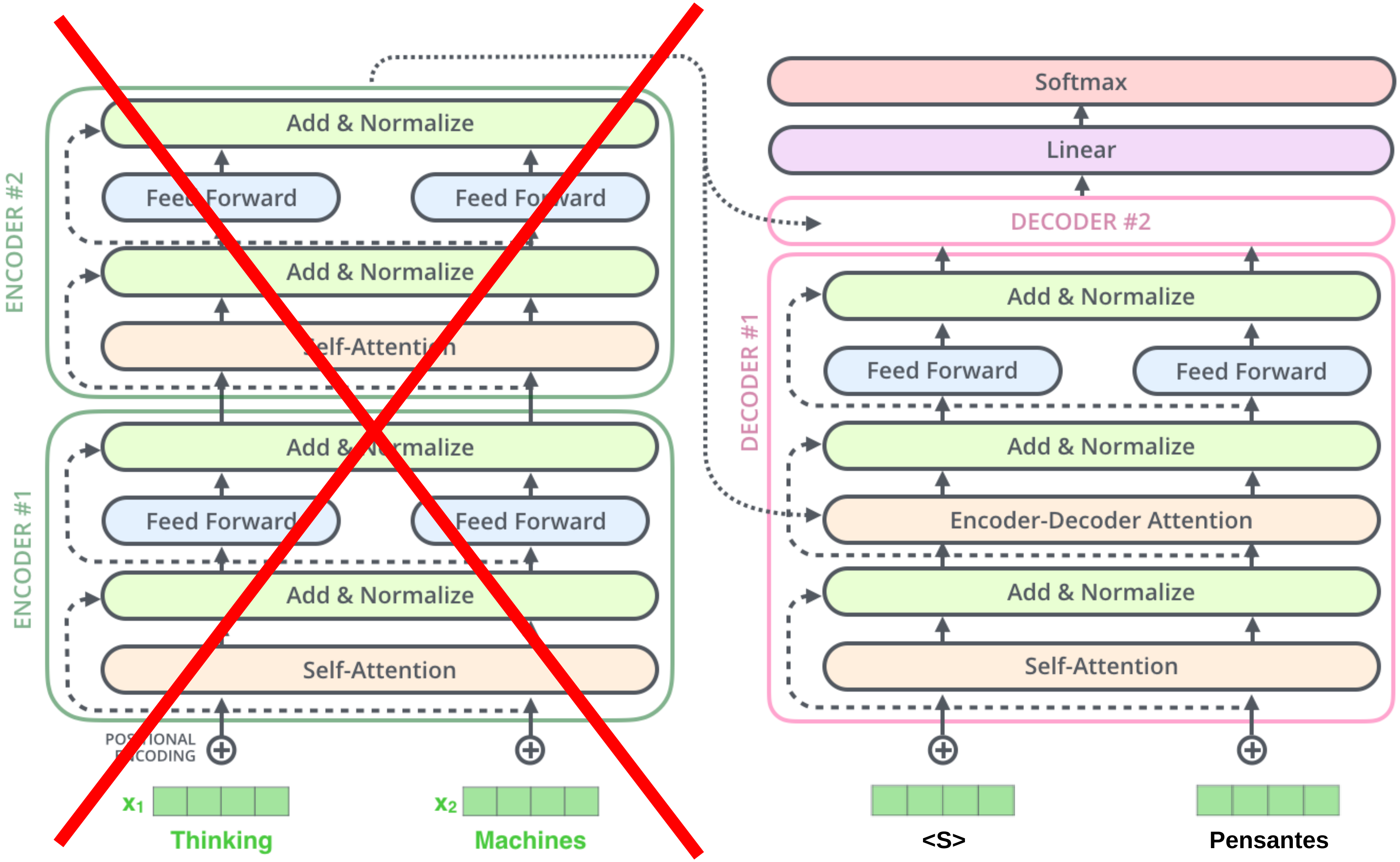
PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

IIC 3800 Tópicos en CC
NLP

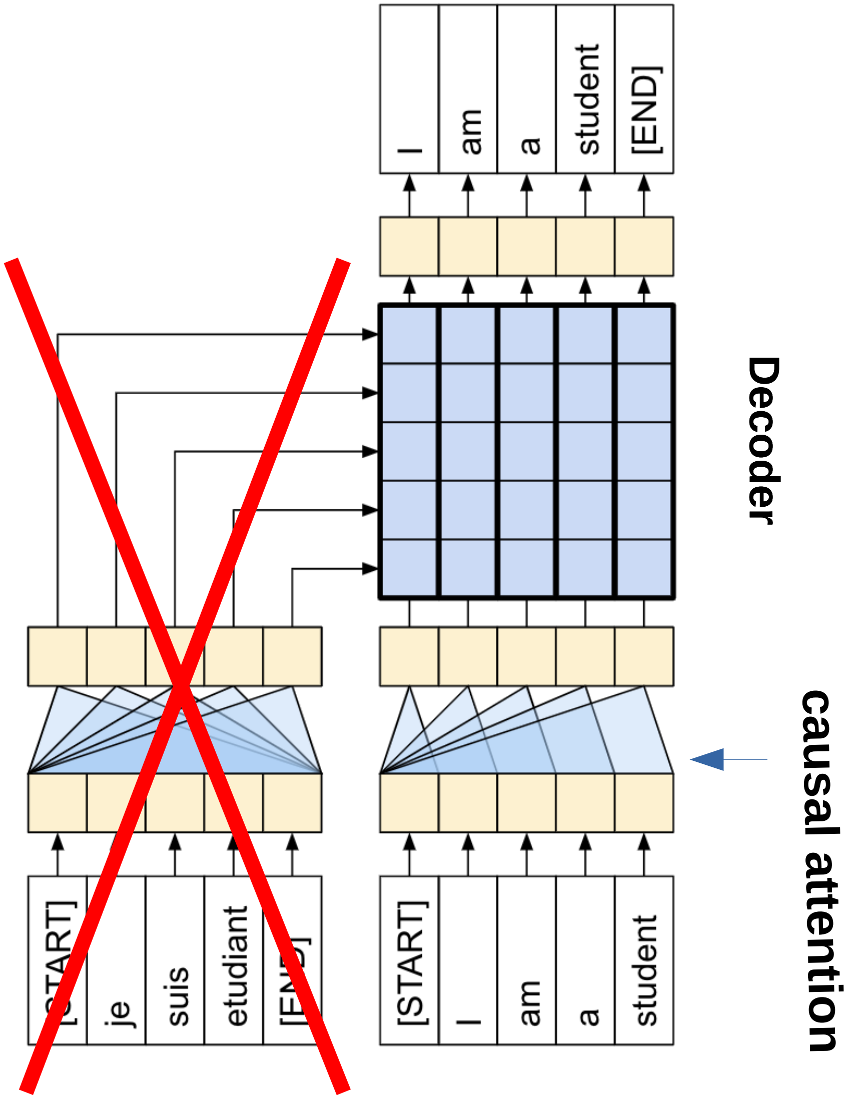
https://github.com/marcelomendoza/IIC3800

- GENERATIVE PRETRAINED TRANSFORMERS -

Se elimina el encoder!!!



Se elimina el encoder!!!



Transformer decoder

Seq2seq

- Cada secuencia $(m^1, ..., m^n) \mapsto (y^1, ..., y^\eta)$

se transforma en:

Símbolo separador

$$(w^1, ..., w^{n+\eta+1}) = (m^1, ..., m^n, \delta, y^1, ..., y^\eta)$$

- Luego, el **decoder** resuelve la siguiente tarea:

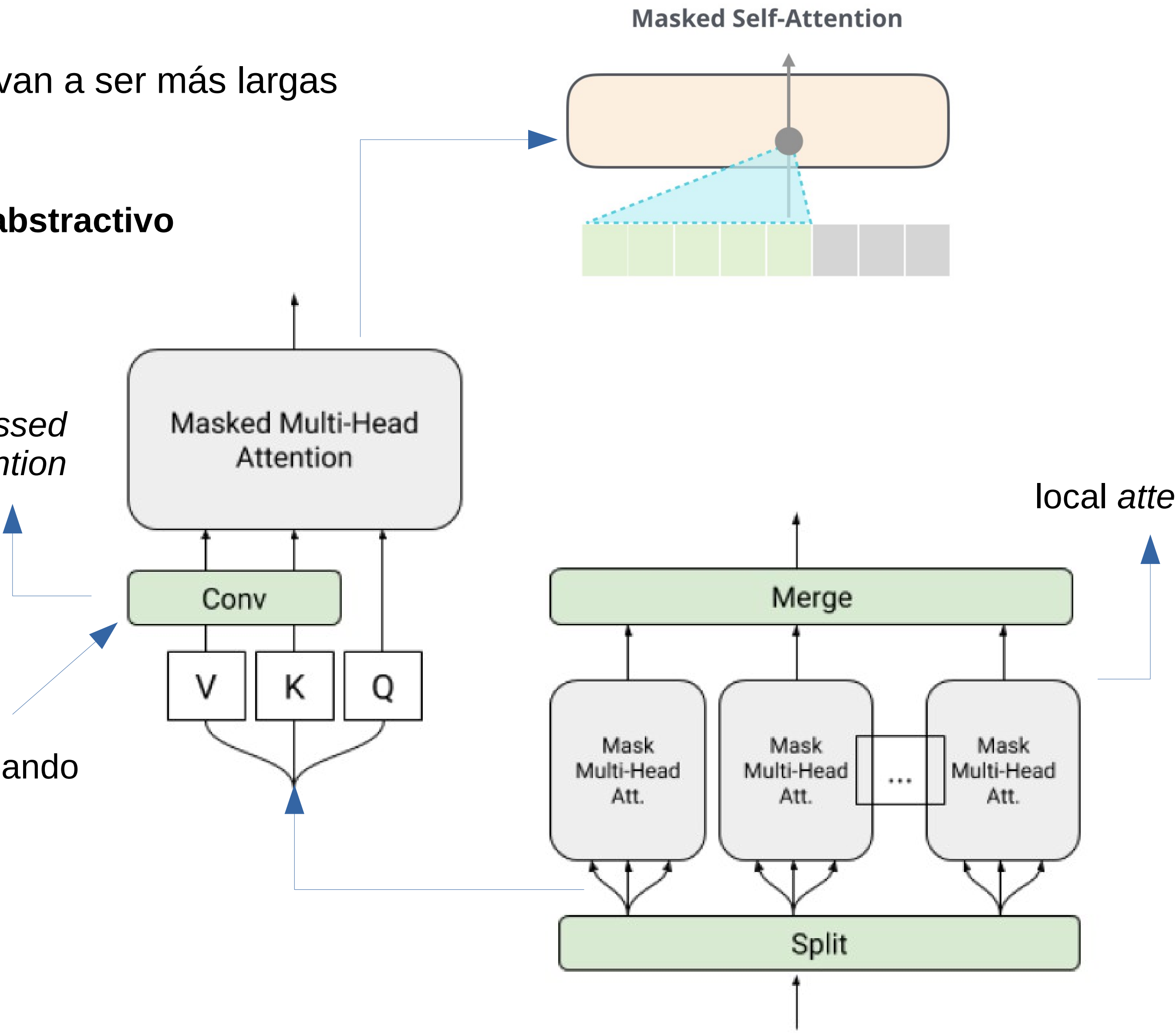
secuencias largas

$$p(w^1, ..., w^{n+\eta}) = \prod_{j=1}^{n+\eta} p(w^i | w^1, ..., w^{j-1})$$

Transformer decoder

Las oraciones van a ser más largas

Modelo **abstractivo**



Memory-compressed attention

Reduce las V y K usando kernels de 3x3

local *attention*

Divide la oración en *batches* del mismo tamaño

Transformer decoder

Mecanismo auto-regresivo

LLM

tokens

$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \ldots, u_{i-1}; \Theta)$$

$$U = (u_{-k}, \ldots, u_{-1})$$

contexto de u

Token embedding matrix

$$h_0 = UW_e + W_p$$

Position embedding matrix

$$h_l = \texttt{transformer\_block}(h_{l-1}) \forall i \in [1, n]$$

#_layers

task

$$P(u) = \texttt{softmax}(h_n W_e^T)$$

Idea propuesta en:

☞ GENERATING WIKIPEDIA BY SUMMARIZING LONG SEQUENCES
Peter J. Liu, Mohammad Saleh,
Etienne Poty, Ben Goodrich, Ryan Sepassi, Łukasz Kaiser, Noam Shazeer
ICLR 2018

GPT 1

Transformer decoder + supervised fine tuning

$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \ldots, u_{i-1}; \Theta)$$

Transformer decoder

$$P(y|x^1, \ldots, x^m) = \texttt{softmax}(h_l^m W_y).$$

Supervised fine-tuning

label

Última capa del transformer decoder

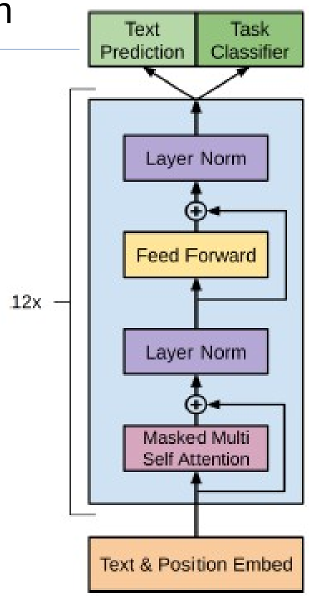
en LM *y* es un token

$$L_2(\mathcal{C}) = \sum_{(x,y)} \log P(y|x^1, \ldots, x^m).$$

La pérdida combinada es:

$$L_3(\mathcal{C}) = L_2(\mathcal{C}) + \lambda * L_1(\mathcal{C})$$

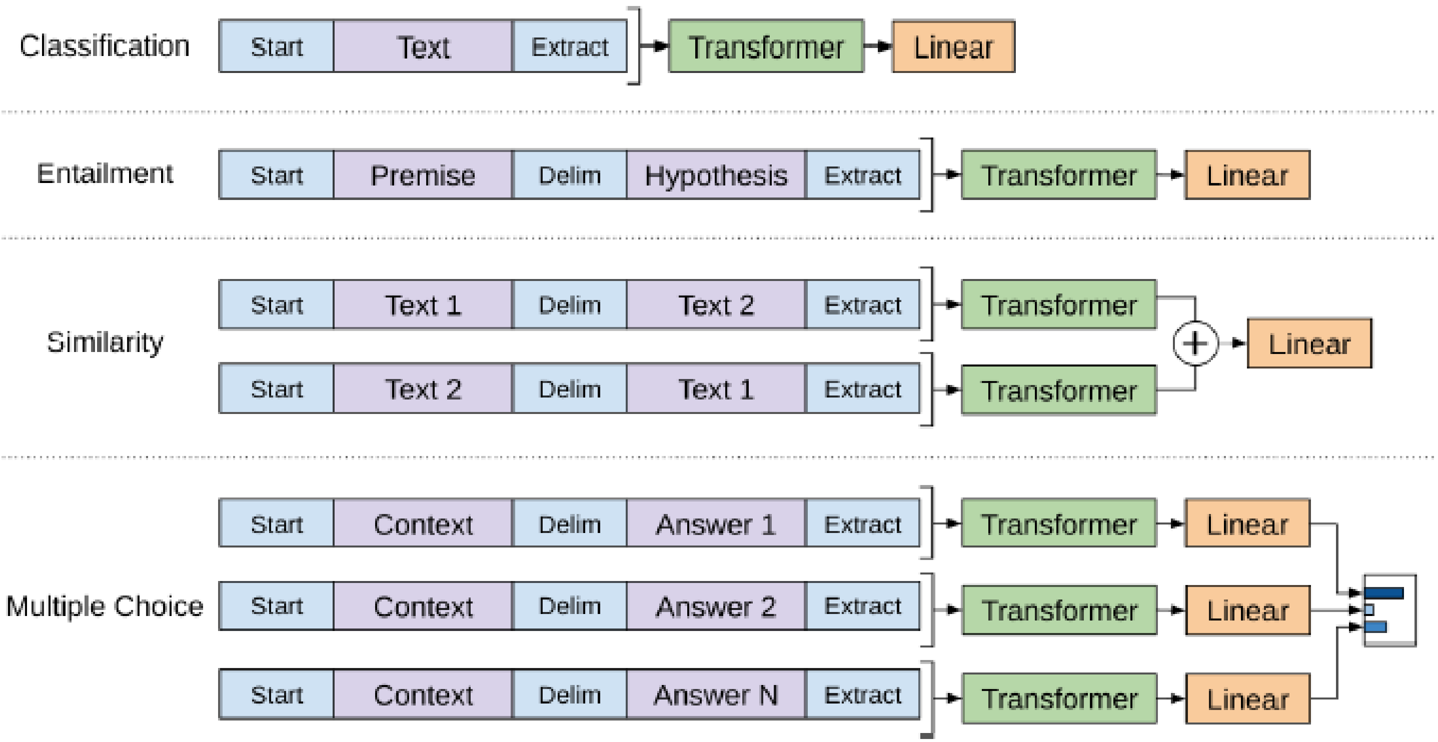
GPT 1 resuelve ambas tareas simultáneamente (LM y specific task)



GPT 1

Supervised fine-tuning

La secuencia de entrada depende de la tarea



OpenAI

GPT 1 ☞ Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. Improving language understanding by generative pre-training, 2018.