Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ingeniería de Transporte y Logística

Sistemas Urbanos Inteligentes

Foundation Models

Hans Löbel
Dpto. Ingeniería de Transporte y Logística
Dpto. Ciencia de la Computación

Center for Research on Foundation Models

Stanford University Human-Centered Artificial Intelligence

On the Opportunities and Risks of Foundation Models

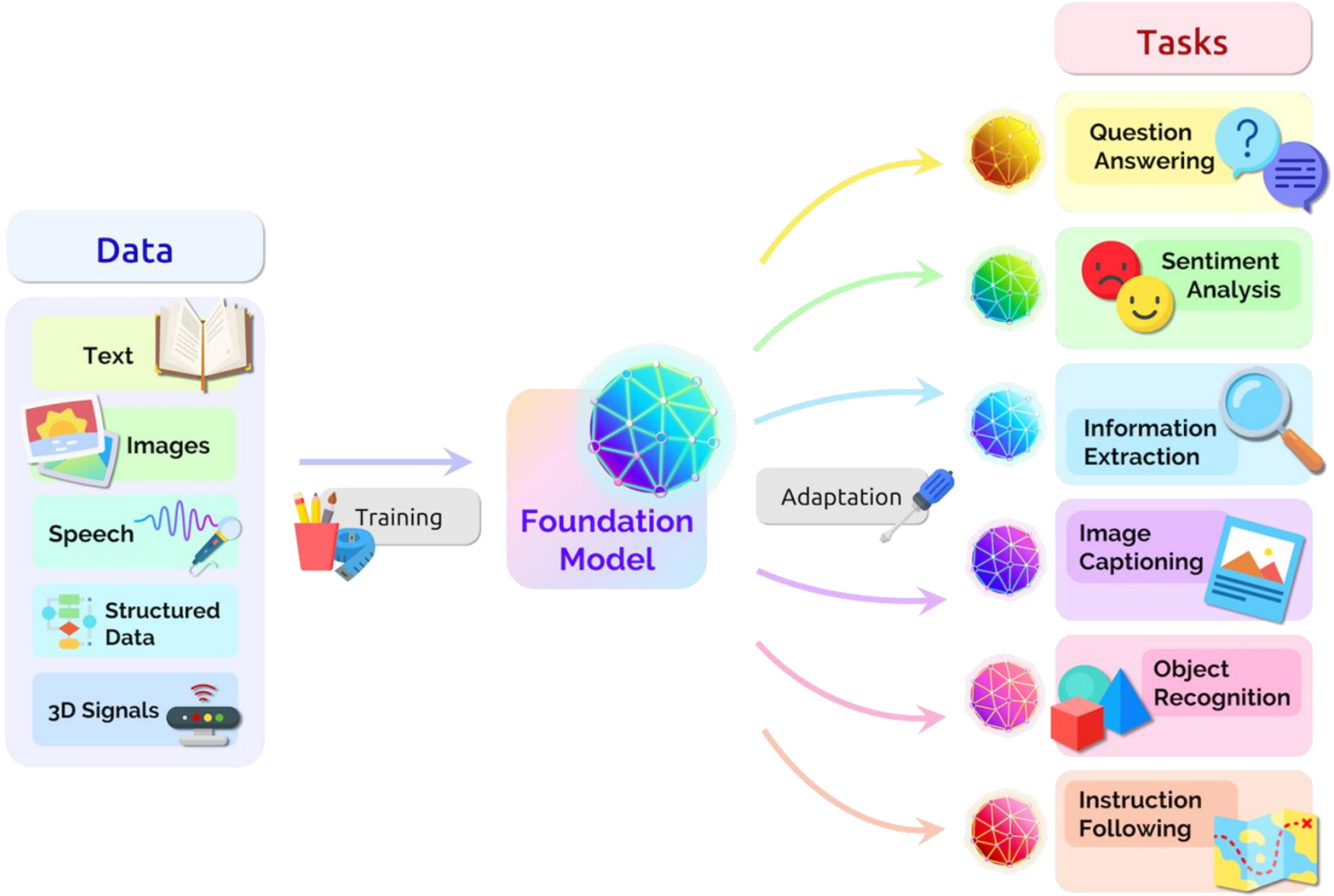
Rishi Bommasani* Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora
Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill
Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji
Annie Chen Kathleen Creel Jared Quincy Davis Dorottya Demszky Chris Donahue
Moussa Doumbouya Esin Durmus Stefano Ermon John Etchemendy Kawin Ethayarajh
Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman
Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt
Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Saahil Jain
Dan Jurafsky Pratyusha Kalluri Siddharth Karamcheti Geoff Keeling Fereshte Khani
Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kuditipudi
Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent
Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning
Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avanika Narayan
Deepak Narayanan Ben Newman Allen Nie Juan Carlos Niebles Hamed Nilforoshan
Julian Nyarko Giray Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech
Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren
Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dorsa Sadigh
Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin
Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu
Jiajun Wu Yuhuai Wu Sang Michael Xie Michihiro Yasunaga Jiaxuan You Matei Zaharia
Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou
Percy Liang*[1]

Center for Research on Foundation Models (CRFM)
Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Stanford University

AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) that are trained on broad data at scale and are adaptable to a wide range of downstream tasks. We call these models foundation models *to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotics, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations). Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization. Homogenization provides powerful leverage but demands caution, as the defects of the foundation model are inherited by all the adapted models downstream. Despite the impending widespread deployment of foundation models, we currently lack a clear understanding of how they work, when they fail, and what they are even capable of due to their emergent properties. To tackle these questions, we believe much of the critical research on foundation models will require deep interdisciplinary collaboration commensurate with their fundamentally sociotechnical nature.*

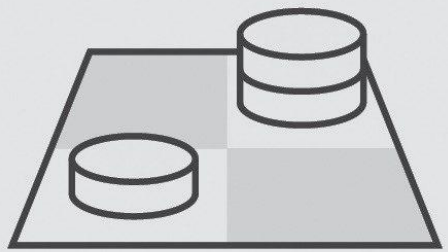
[1]Corresponding author: pliang@cs.stanford.edu

*Equal contribution.


Data
Text
Images
Speech
Structured Data
3D Signals
Training
Foundation Model
Adaptation
Tasks
Question Answering
Sentiment Analysis
Information Extraction
Image Captioning
Object Recognition
Instruction Following

ARTIFICIAL INTELLIGENCE

Artificial Intelligence captures the imagination of the world.



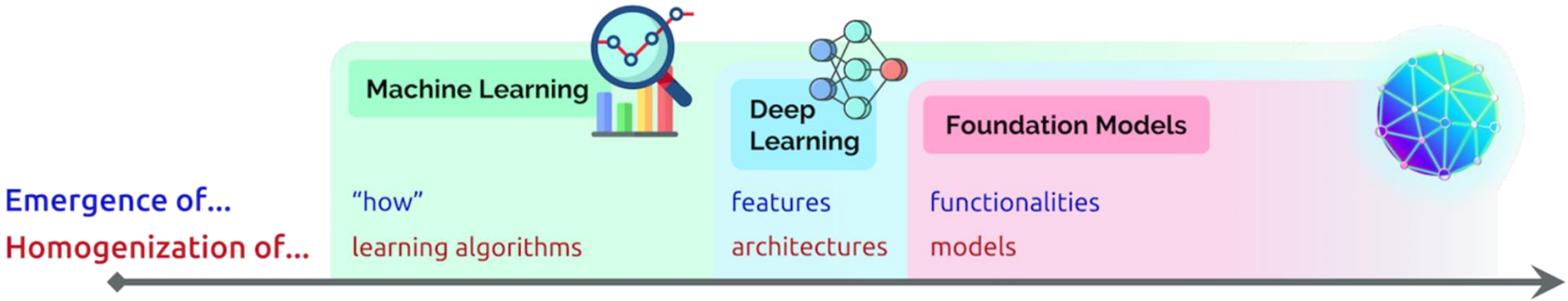
MACHINE LEARNING

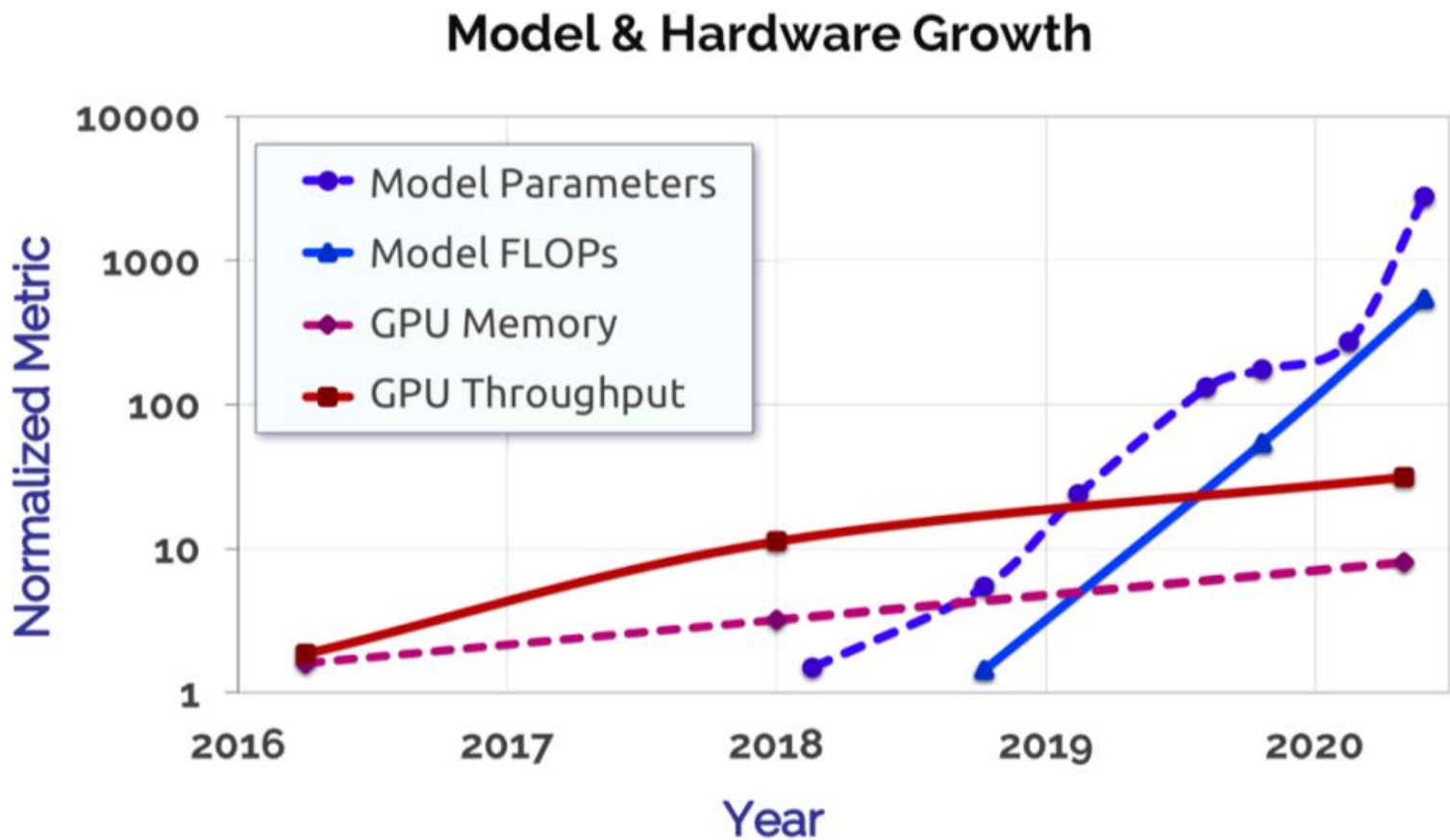
Machine learning starts to gain traction.

DEEP LEARNING

Deep learning catapults the industry.

1950s 1960s 1970s 1980s 1990s 2000s 2010s


Machine Learning
Deep Learning
Foundation Models
Emergence of...
Homogenization of...
"how"
learning algorithms
features
architectures
functionalities
models


Model & Hardware Growth
10000
1000
100
10
1
Normalized Metric
Model Parameters
Model FLOPs
GPU Memory
GPU Throughput
2016
2017
2018
2019
2020
Year

Aplicaciones multimodales: texto e imágenes (DALL-E 2)

DALL·E 2 can create original, realistic images and art from a text description. It can combine concepts, attributes, and styles.

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

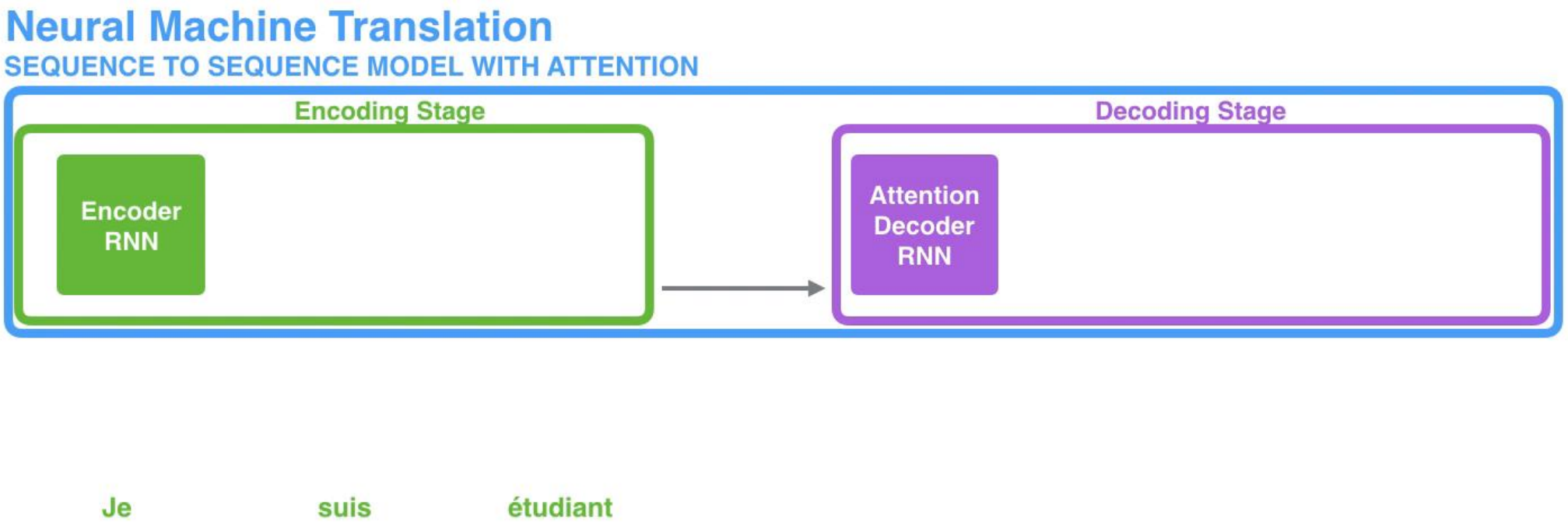
riding a horse lounging in a tropical resort in space playing basketball with cats in space

in a vaporwave style as pixel art in a photorealistic style

→

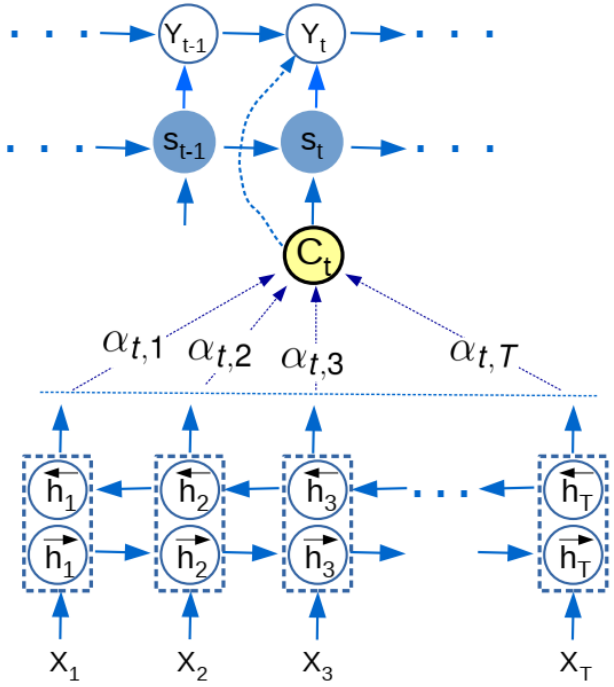
DALL·E 2

Modelos seq2seq con atención conceptualmente funcionan muy bien

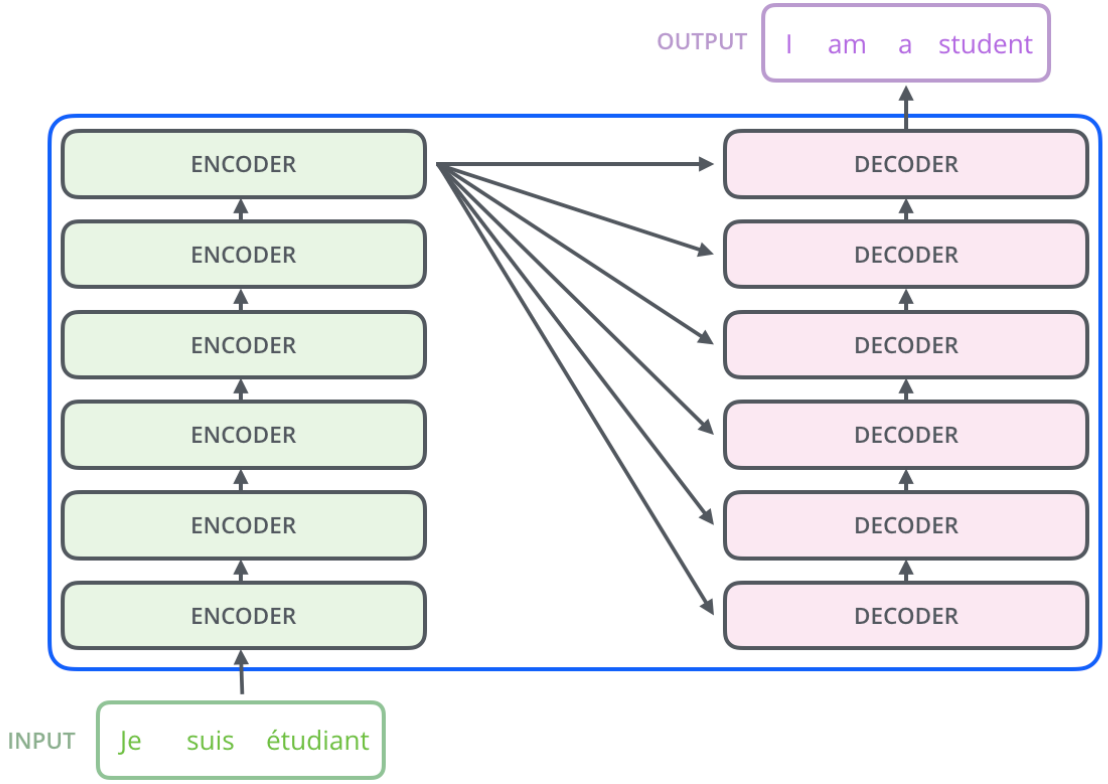


Lamentablemente, acarrean muchos de los problemas de las RNN

- Poco eficientes computacionalmente.
- Problemas con secuencias muy largas.
- Estos problemas complican su aplicación a sets de datos gigantescos, que potencialmente entregan mayor conocimiento

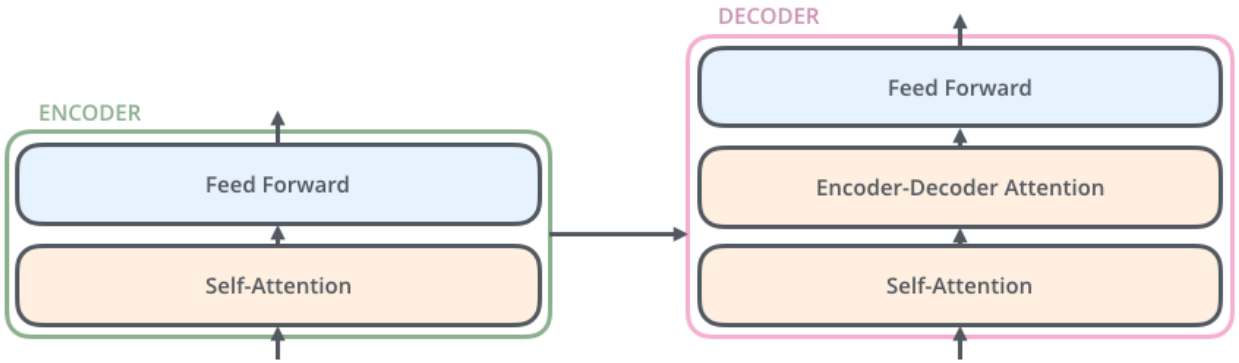


La solución más popular a estos problemas la entrega la arquitectura Transformer



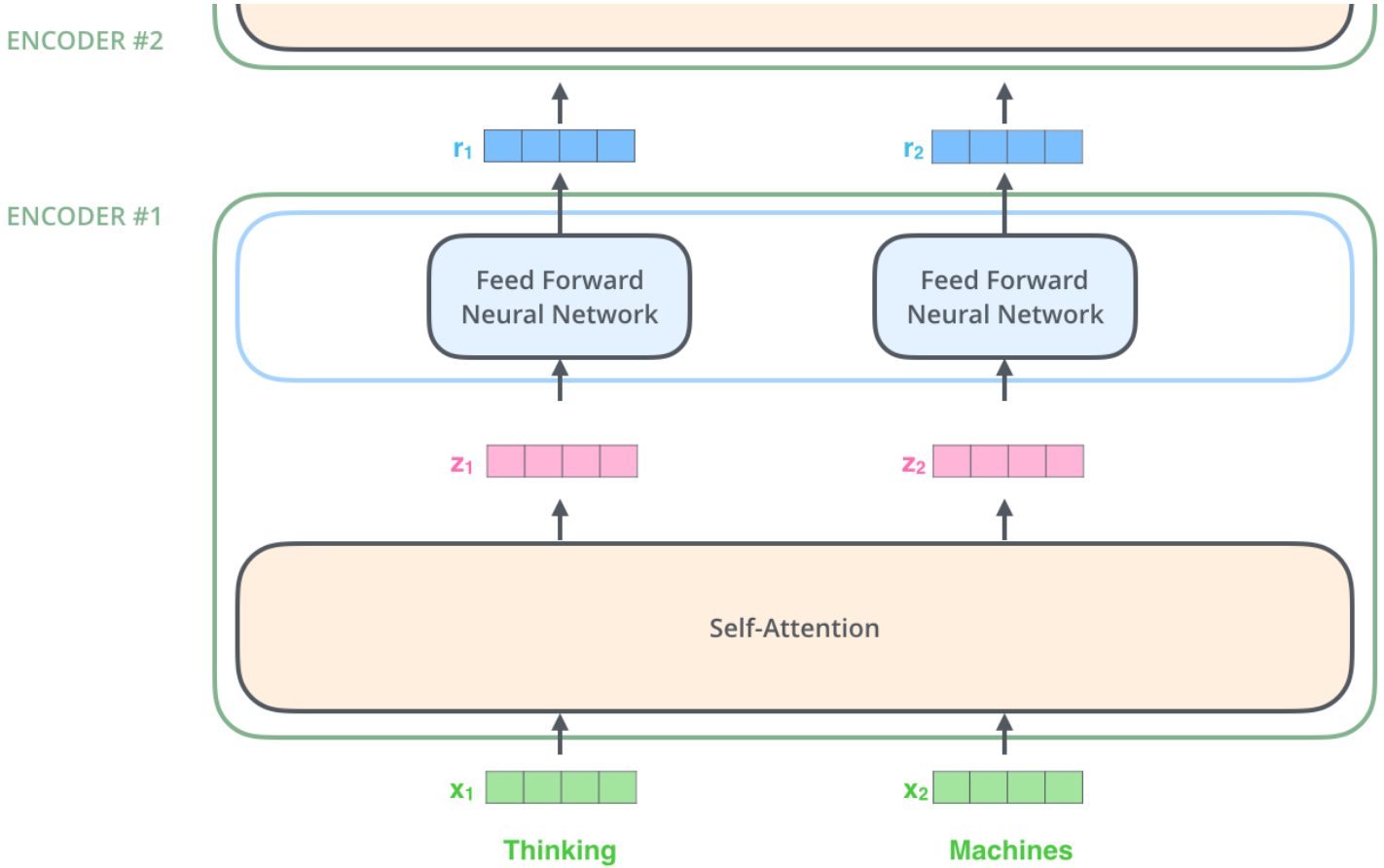
- Esta arquitectura profunda está completamente basada en mecanismos de atención.
- Su gran aporte es ser más eficiente y permitir dependencias de mayor largo que los modelos seq2seq.

Si bien también están formados por *encoder* y *decoders*, estos no son recurrentes, sino combinaciones de atención y capas densas

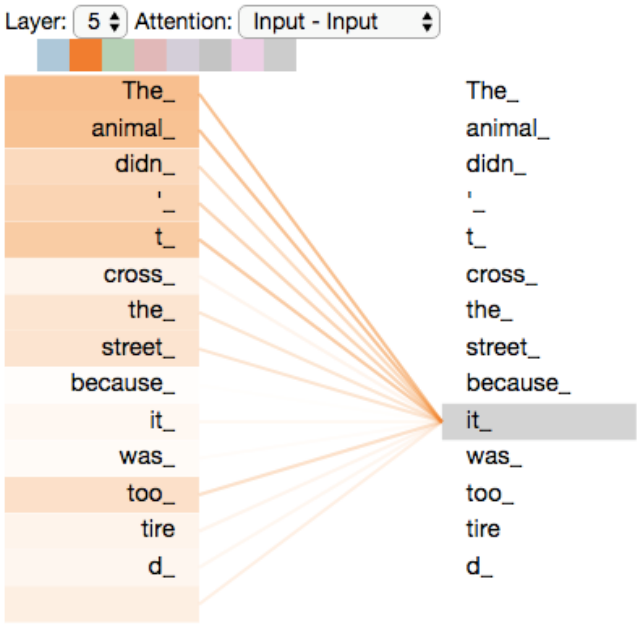


- Atención en Transformers no es igual a la de un modelo seq2seq.
- En este caso se utiliza la auto-atención, que indica para cada elementos de una secuencia, su dependencia con otros elementos de la misma.

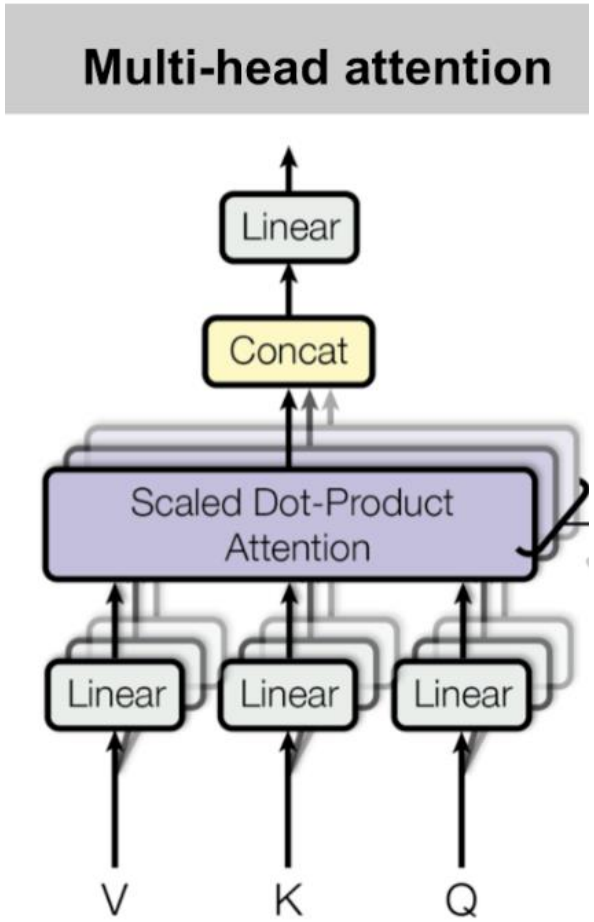
Si bien también están formados por *encoder* y *decoders*, estos no son recurrentes, sino combinaciones de atención y capas densas



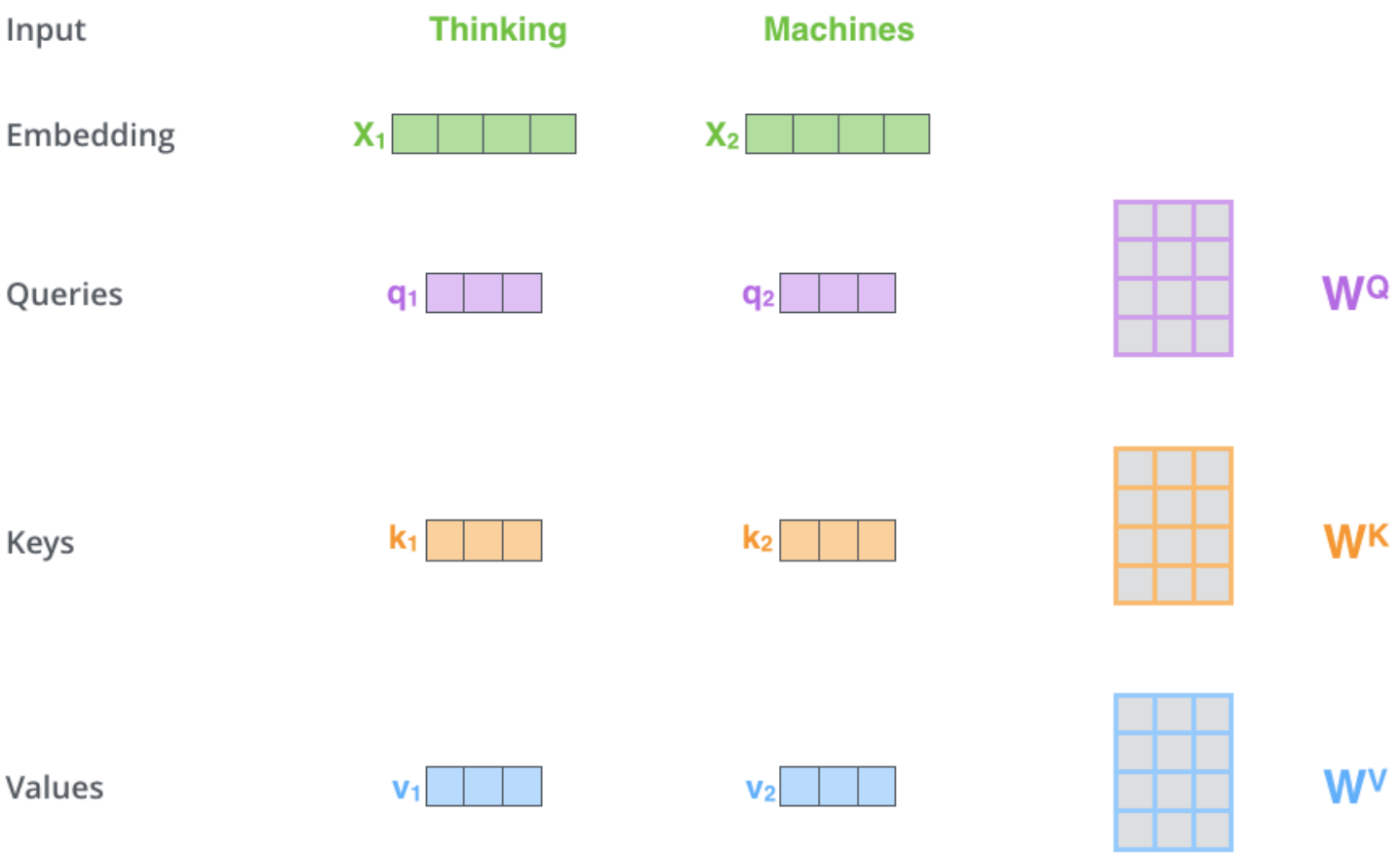
Auto-atención resulta ser el elemento clave



Auto-atención resulta ser el elemento clave

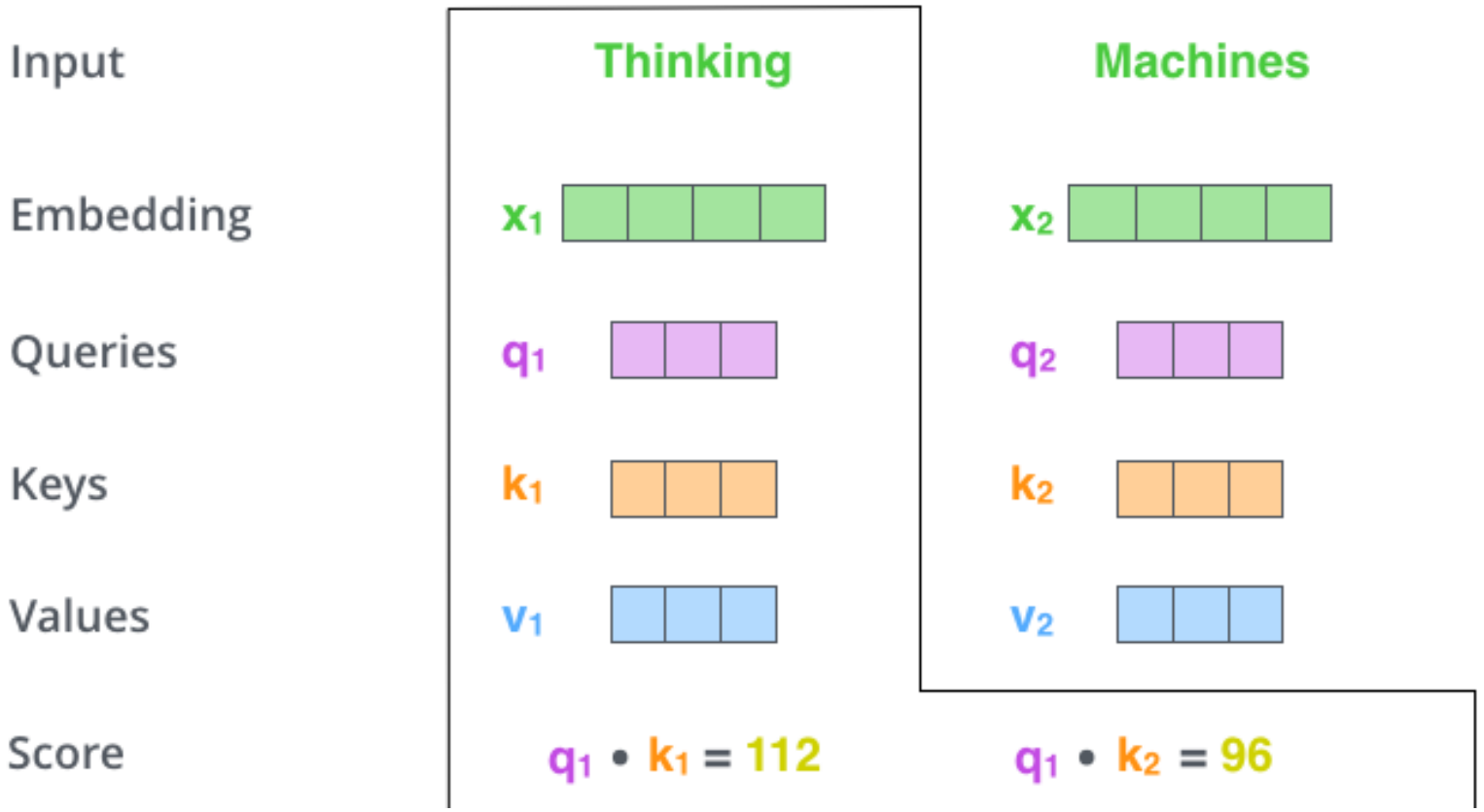


Auto-atención resulta ser el elemento clave



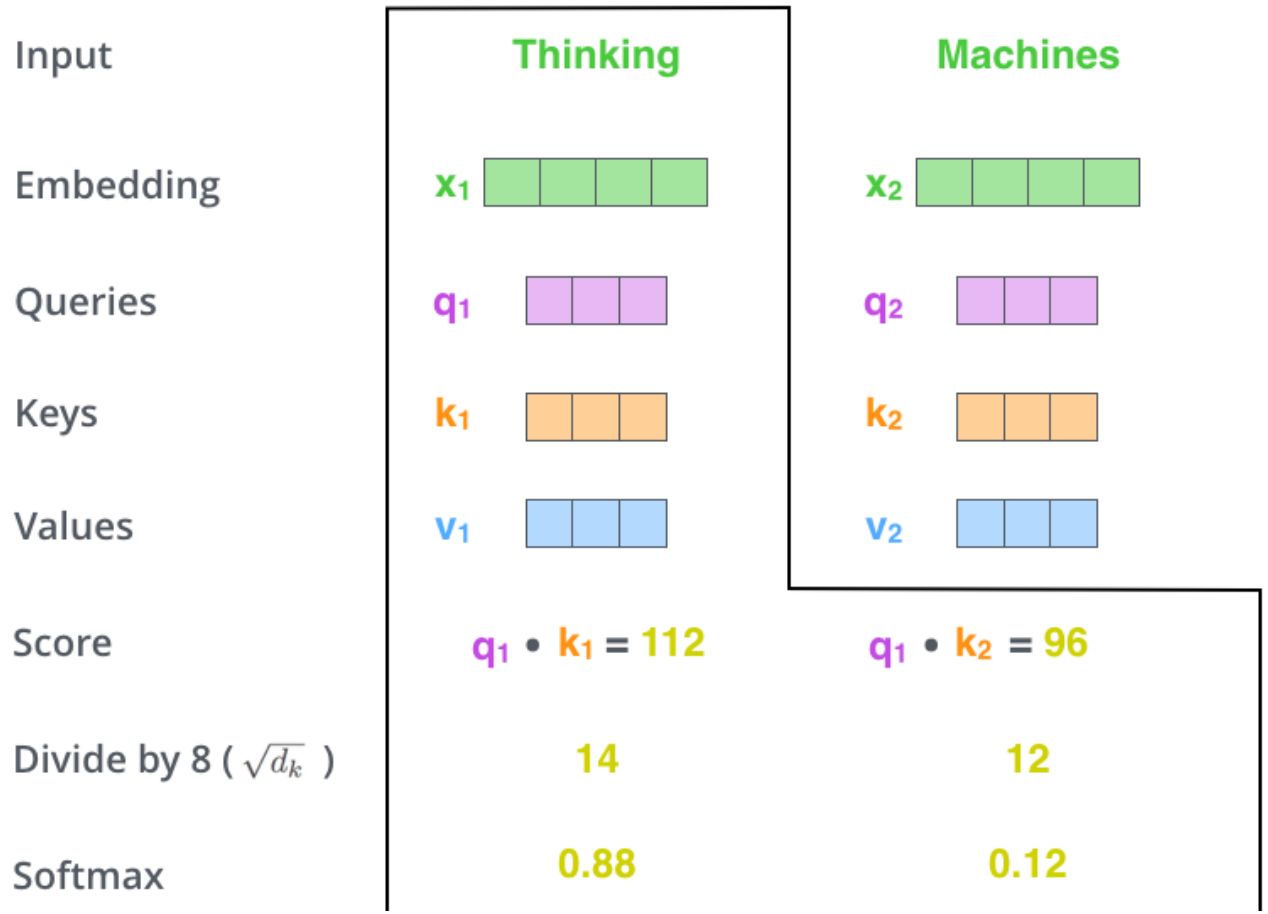
https://jalammar.github.io/illustrated-transformer/

Auto-atención resulta ser el elemento clave



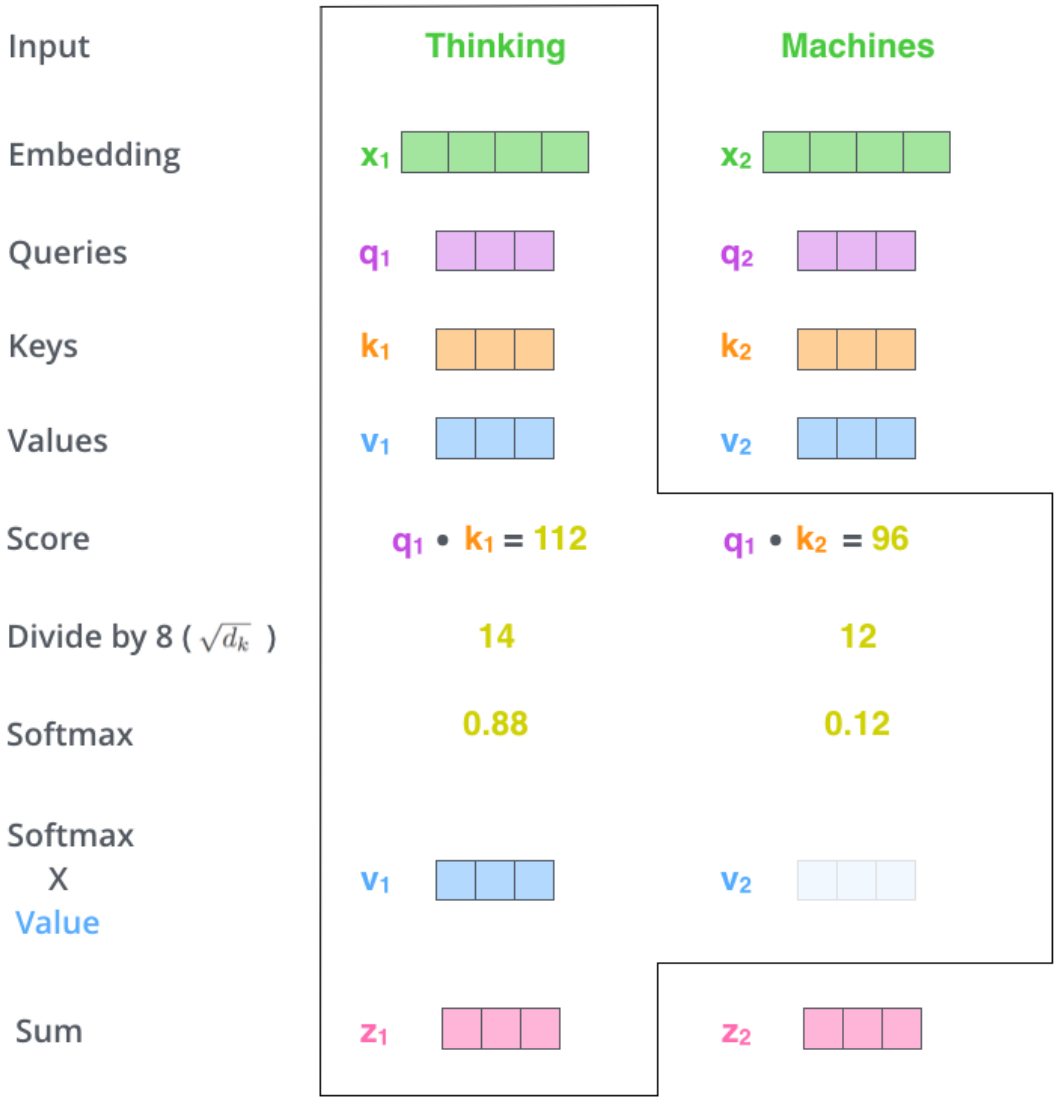
https://jalammar.github.io/illustrated-transformer/

Auto-atención resulta ser el elemento clave



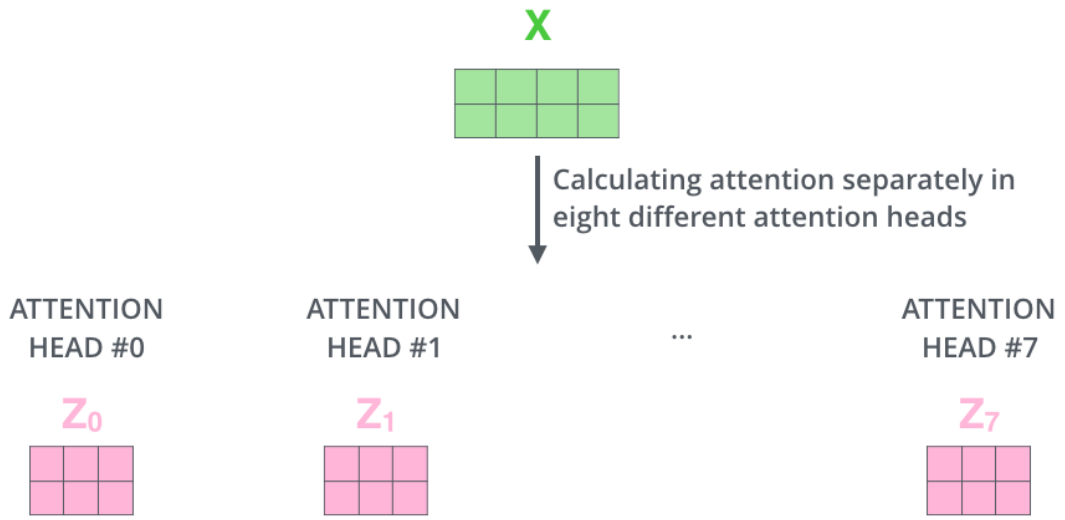
https://jalammar.github.io/illustrated-transformer/

Auto-atención resulta ser el elemento clave



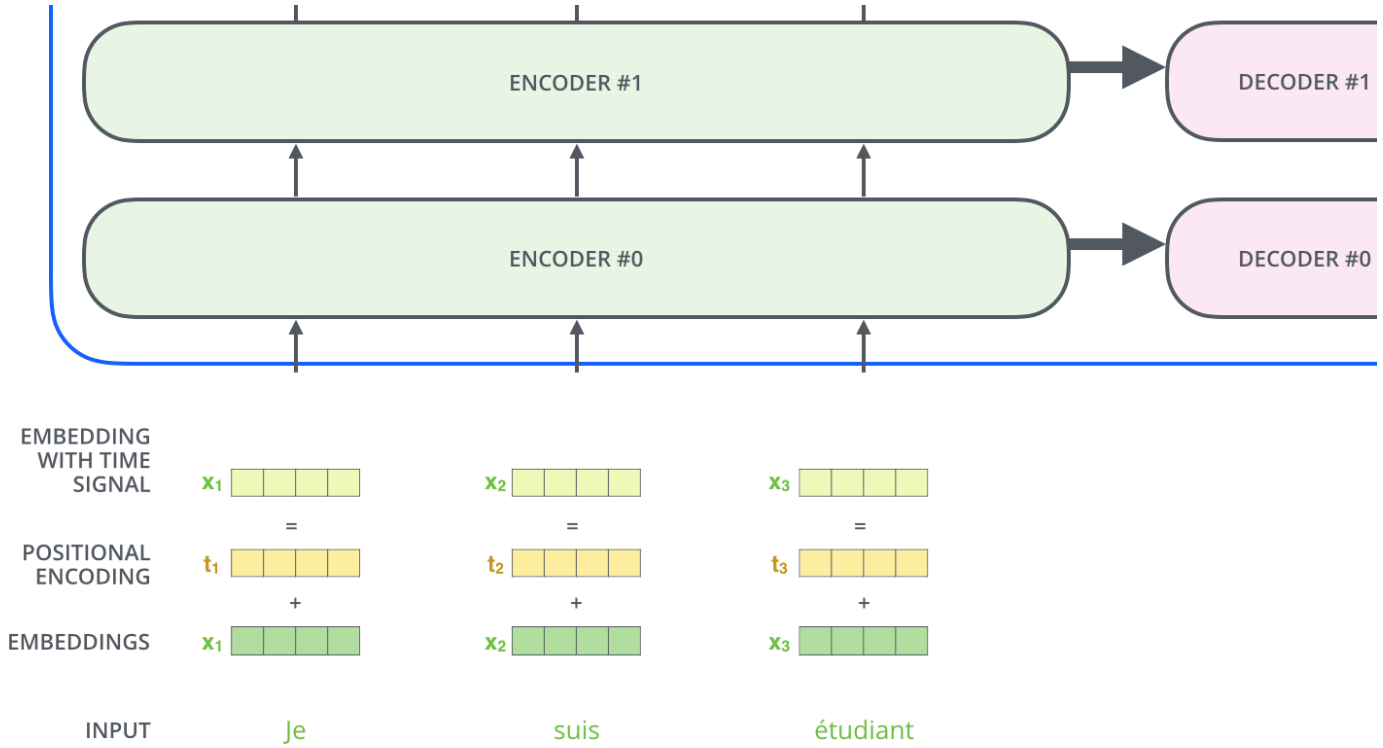
https://jalammar.github.io/illustrated-transformer/

Auto-atención puede ser multimodal (muchas atenciones distintas)



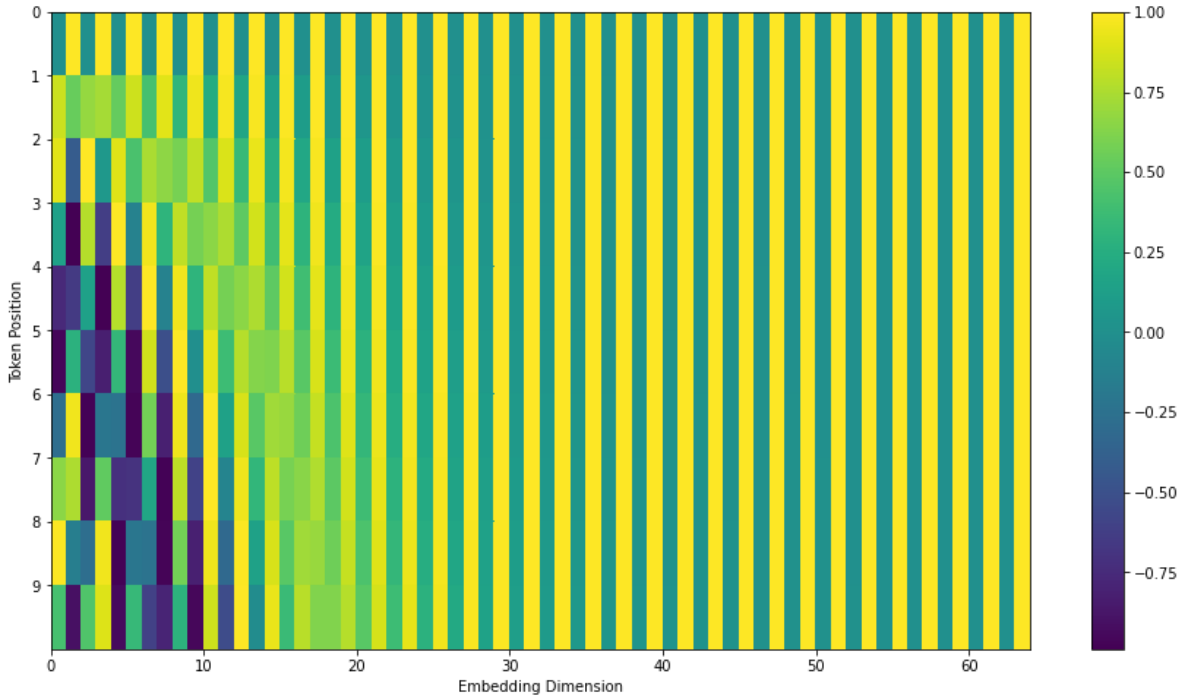
https://jalammar.github.io/illustrated-transformer/

Orden de las secuencias es incluye a través de un *embedding* temporal



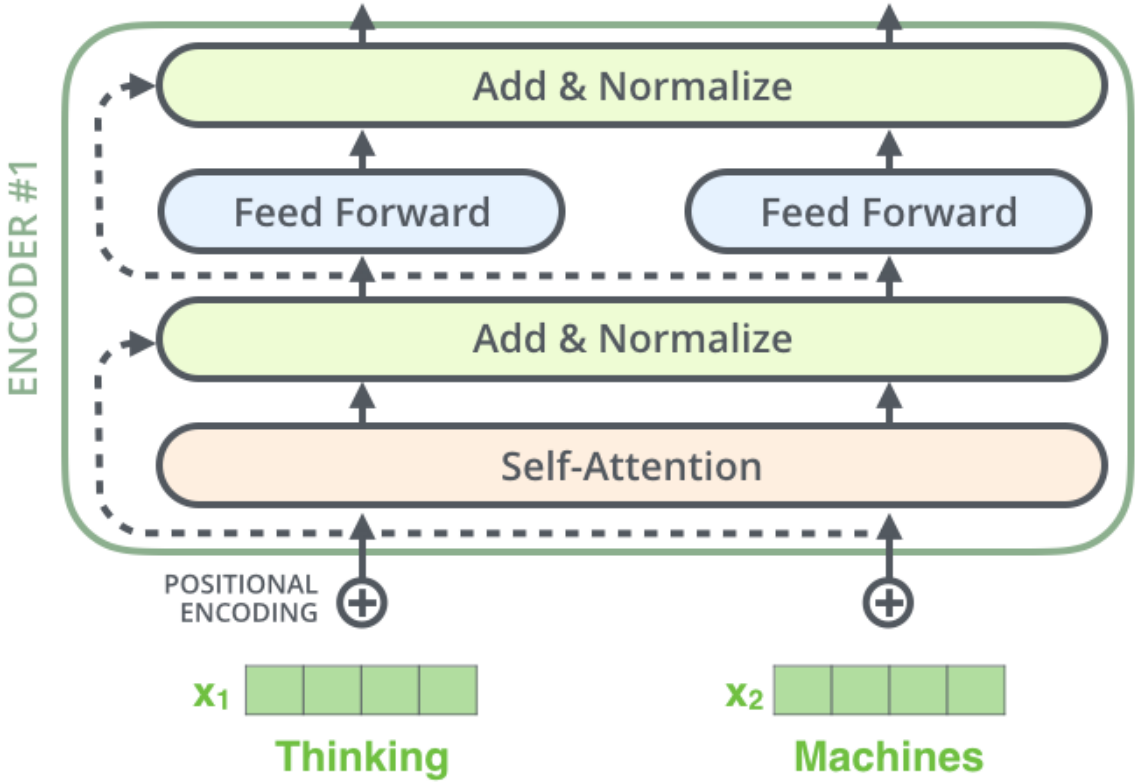
https://jalammar.github.io/illustrated-transformer/

Orden de las secuencias es incluye a través de un *embedding* temporal



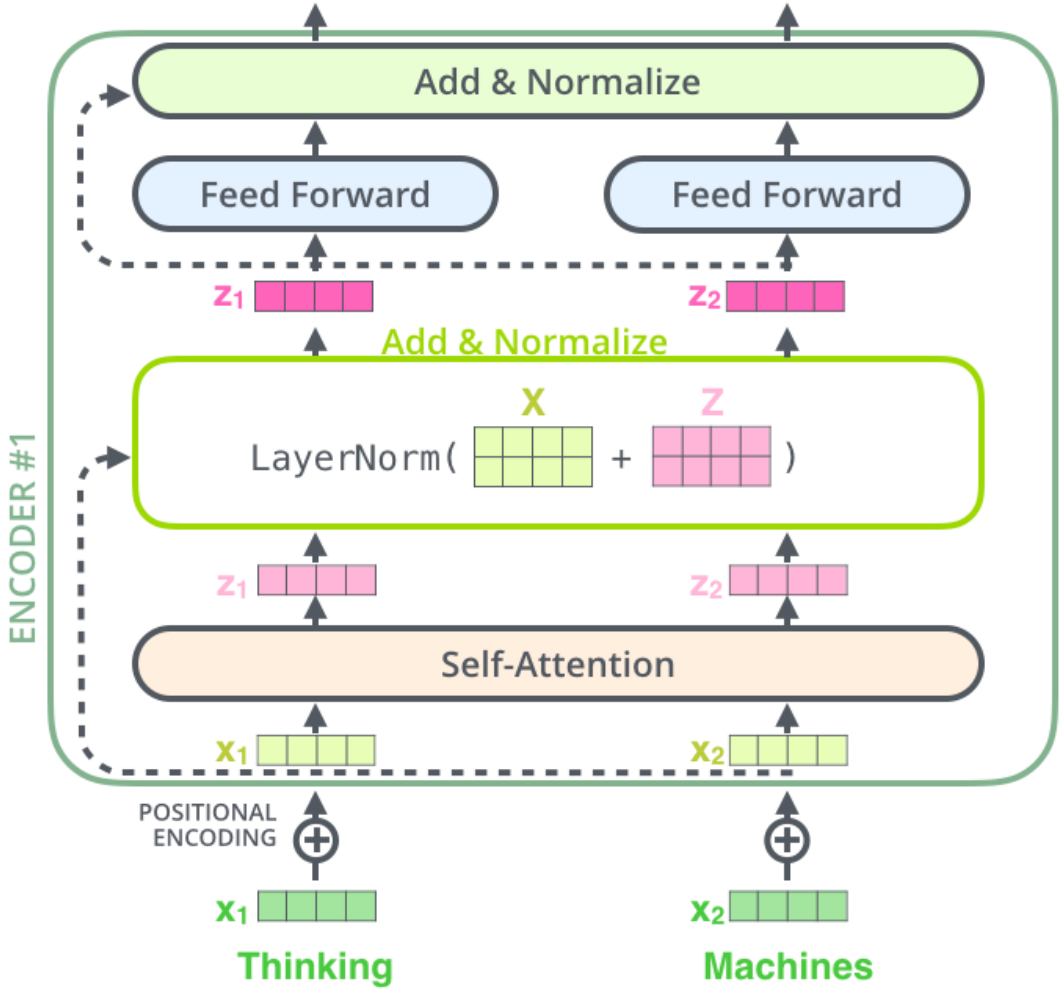
https://jalammar.github.io/illustrated-transformer/

Algunos detalles faltantes: conexiones residuales y normalización



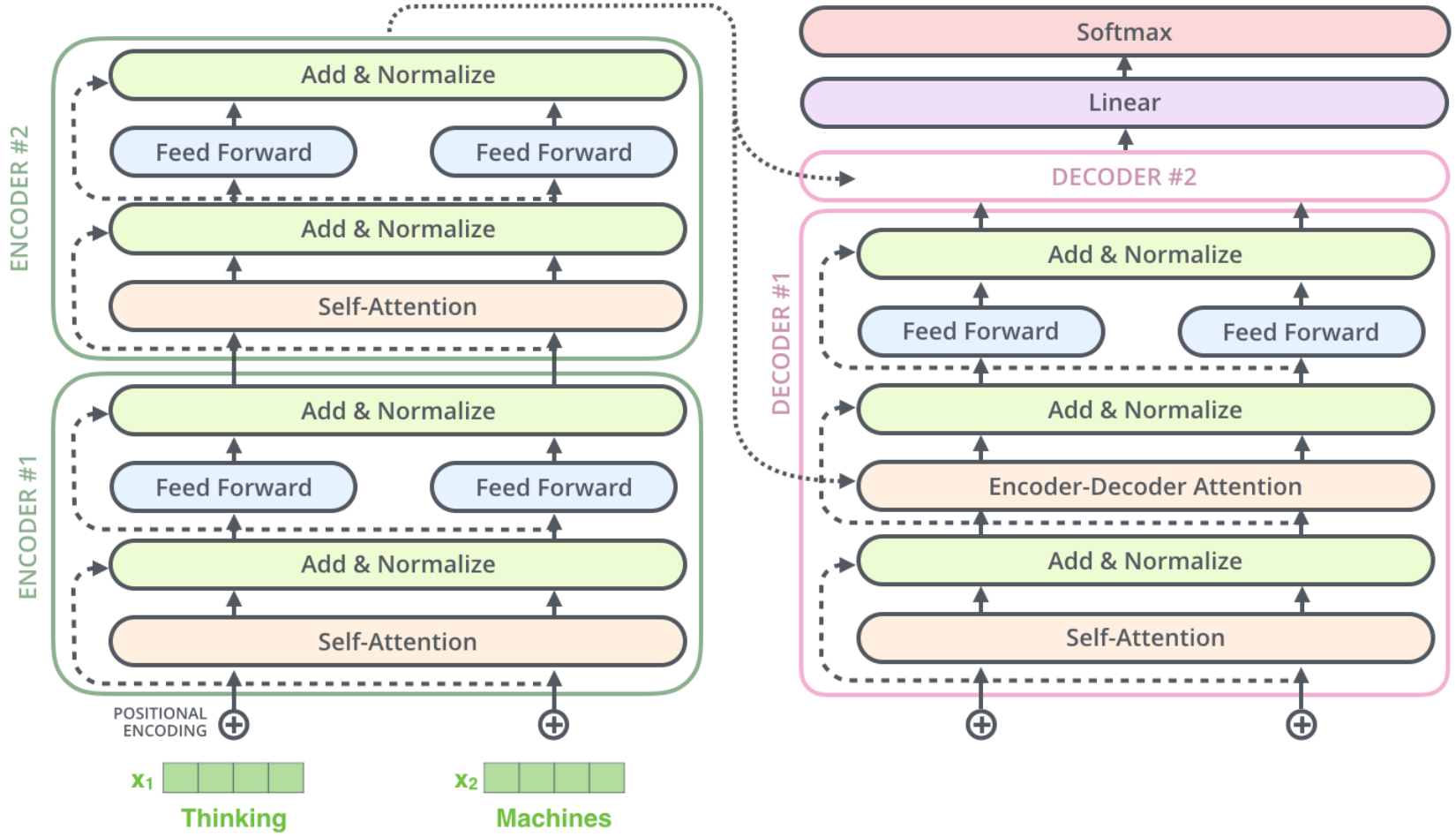
https://jalammar.github.io/illustrated-transformer/

Algunos detalles faltantes: conexiones residuales y normalización



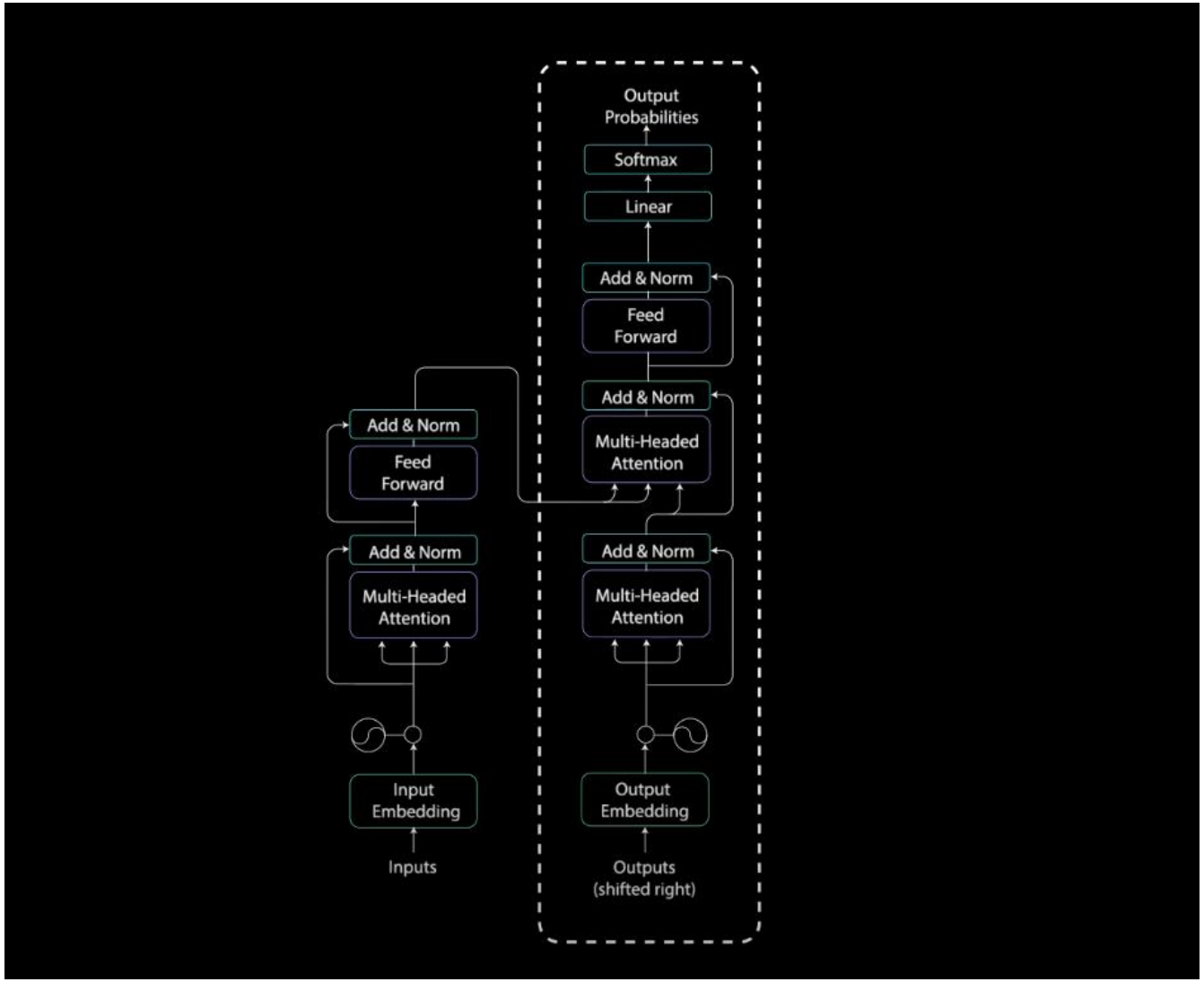
https://jalammar.github.io/illustrated-transformer/

Finalmente, el *decoder* genera las predicciones

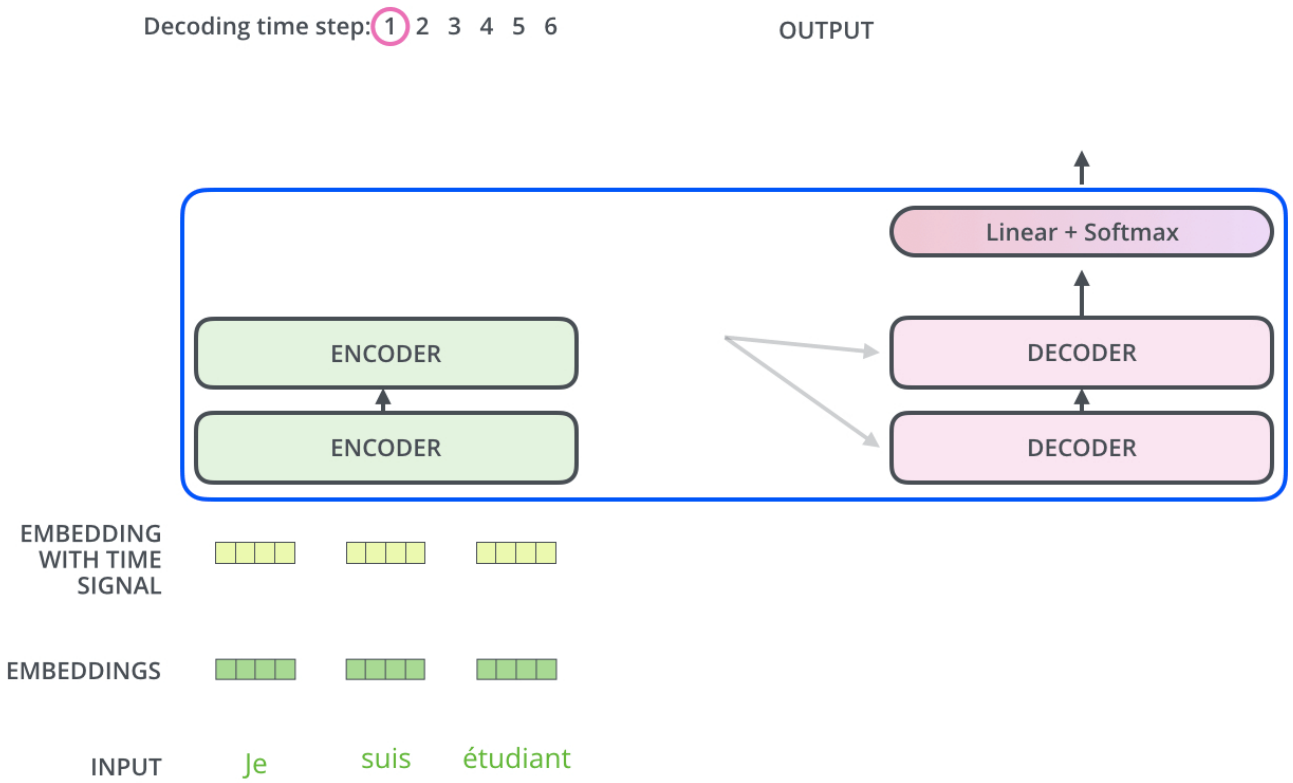


https://jalammar.github.io/illustrated-transformer/

Finalmente, el *decoder* genera las predicciones

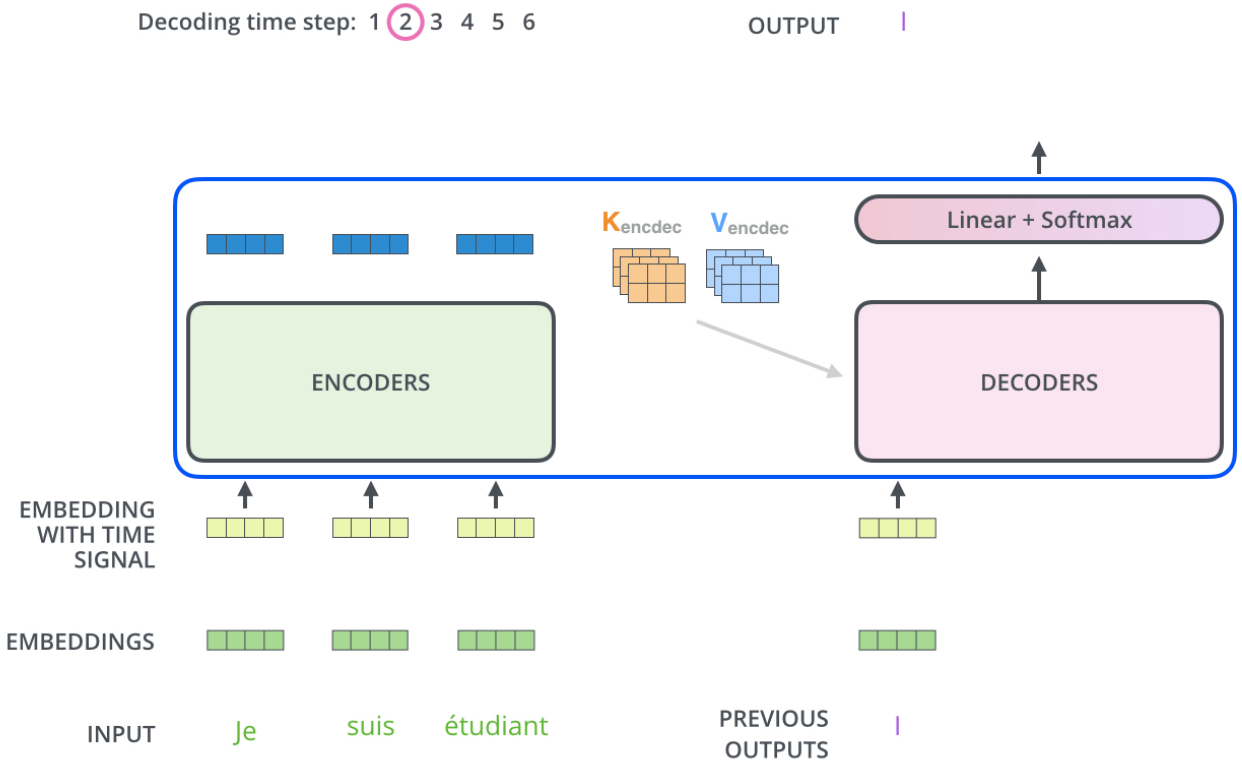


Finalmente, el *decoder* genera las predicciones



https://jalammar.github.io/illustrated-transformer/

Finalmente, el *decoder* genera las predicciones



https://jalammar.github.io/illustrated-transformer/

Transformer fue, indirectamente, el primer paso para el cambio de paradigma actual

- Si bien la arquitectura fue muy novedosa, su proceso de entrenamiento es tradicional.
- Dicho de otra forma, se entrena con una tarea supervisada (*translation*), y luego se hace fine-tuning para llevar el conocimiento a otras tareas.
- Esto es problemático, ya que los datos rotulados no son suficientes para sacar todo el provecho necesario de este modelo.
- El gran gran paso fue como aprovechar esta poderosa arquitectura con un esquema de entrenamiento *adhoc*.

Bert fue uno de los primeros que capitalizó sobre esta idea

1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

Semi-supervised Learning Step

Model:

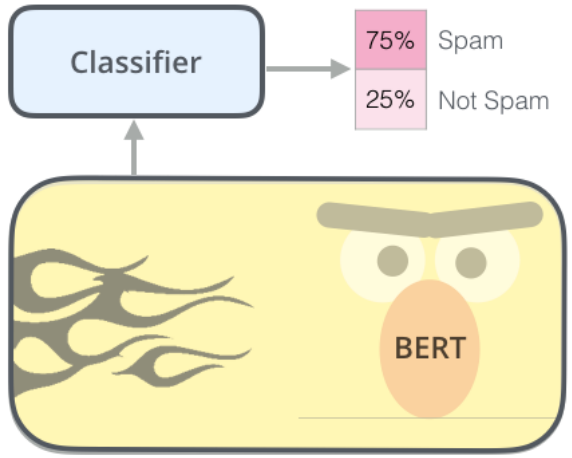
Dataset:

Objective:

Predict the masked word (langauge modeling)

2 - Supervised training on a specific task with a labeled dataset.

Supervised Learning Step



Dataset:

Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached…	Not Spam

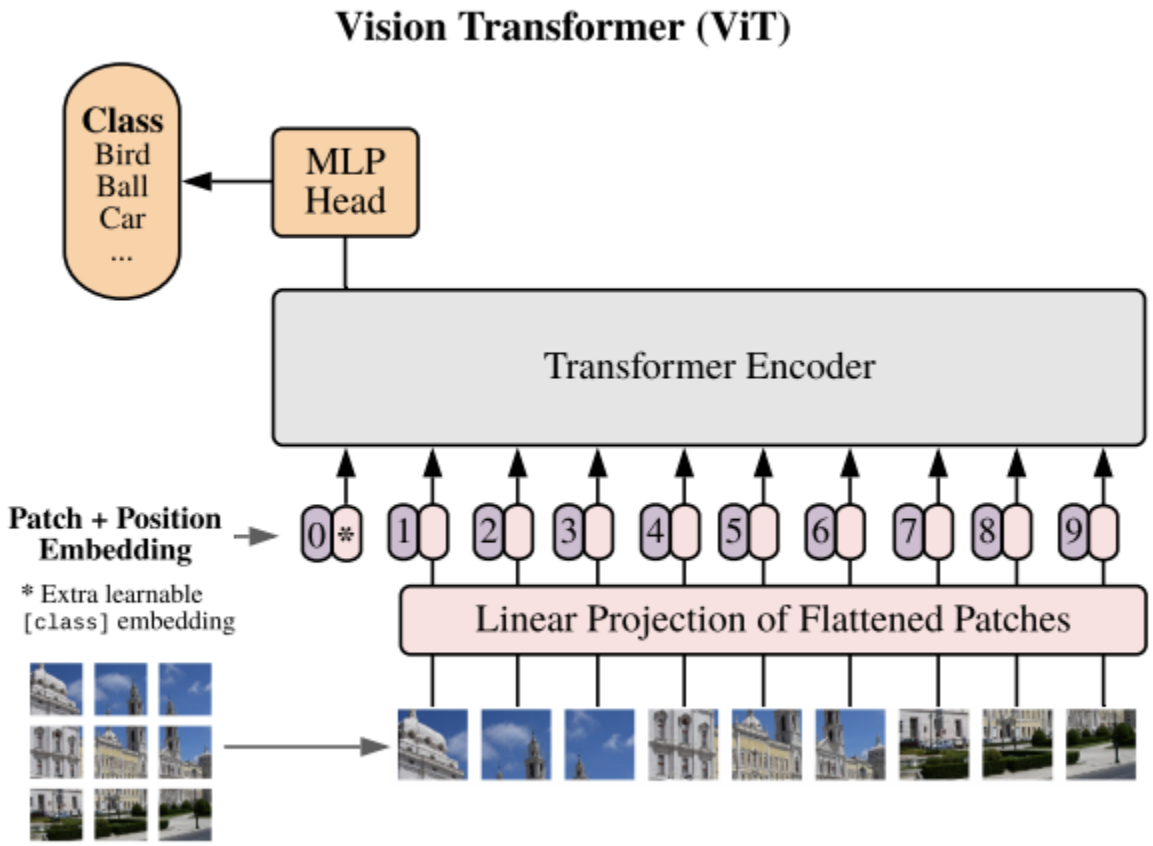
Beto es esencialmente Bert en español

Beto: Bert's model trained with a Spanich corpus.

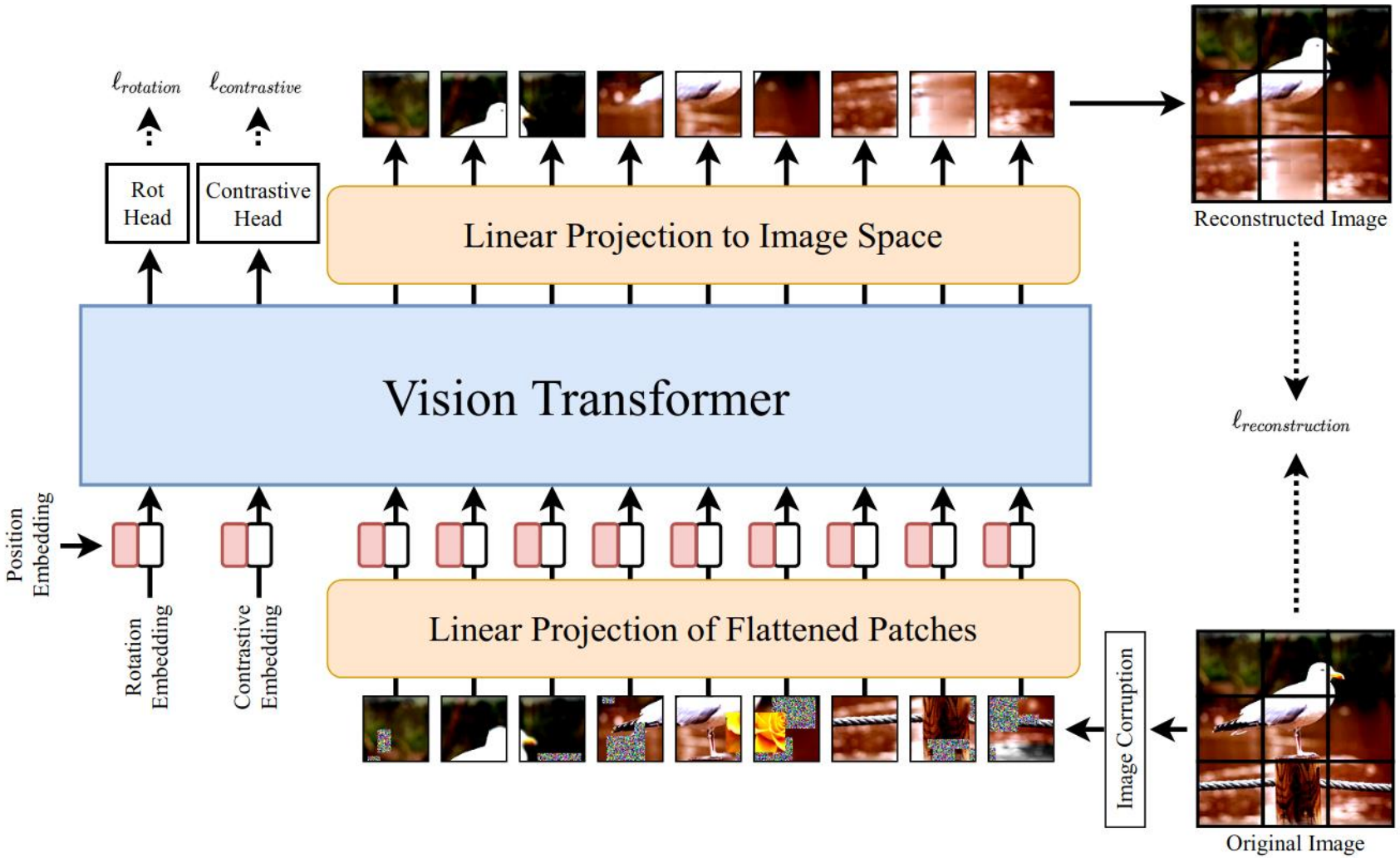
Task	BETO-cased	BETO-uncased	Best Multilingual BERT
POS	**98.97**	98.44	97.10 [2]
NER-C	**88.43**	82.67	87.38 [2]
MLDoc	95.60	**96.12**	95.70 [2]
PAWS-X	89.05	89.55	90.70 [8]
XNLI	**82.01**	80.15	78.50 [2]

https://github.com/dccuchile/beto

Esto también a utilizado en visión, con el Vision Transformer



ViT se ha entrenado incluso de forma auto-supervisada



Para cerrar, veamos a muy alto nivel la arquitectura de DALL-E 2

DALL·E 2 can create original, realistic images and art from a text description. It can combine concepts, attributes, and styles.

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

riding a horse lounging in a tropical resort in space playing basketball with cats in space

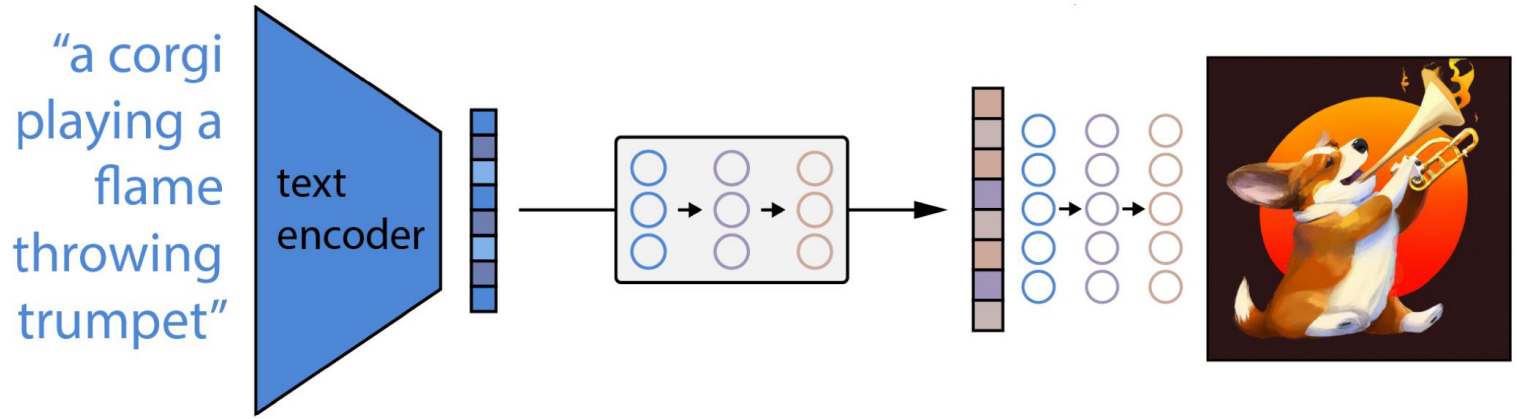
in a vaporwave style as pixel art in a photorealistic style

→

DALL·E 2

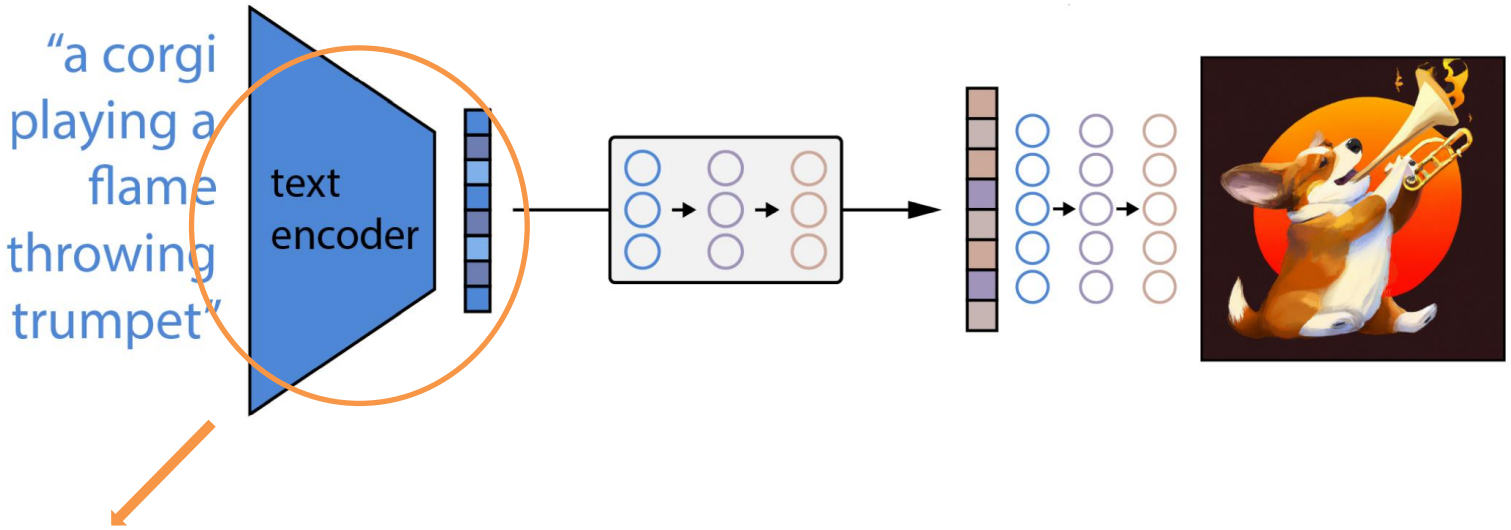
https://openai.com/dall-e-2/

Para cerrar, veamos a muy alto nivel la arquitectura de DALL-E 2



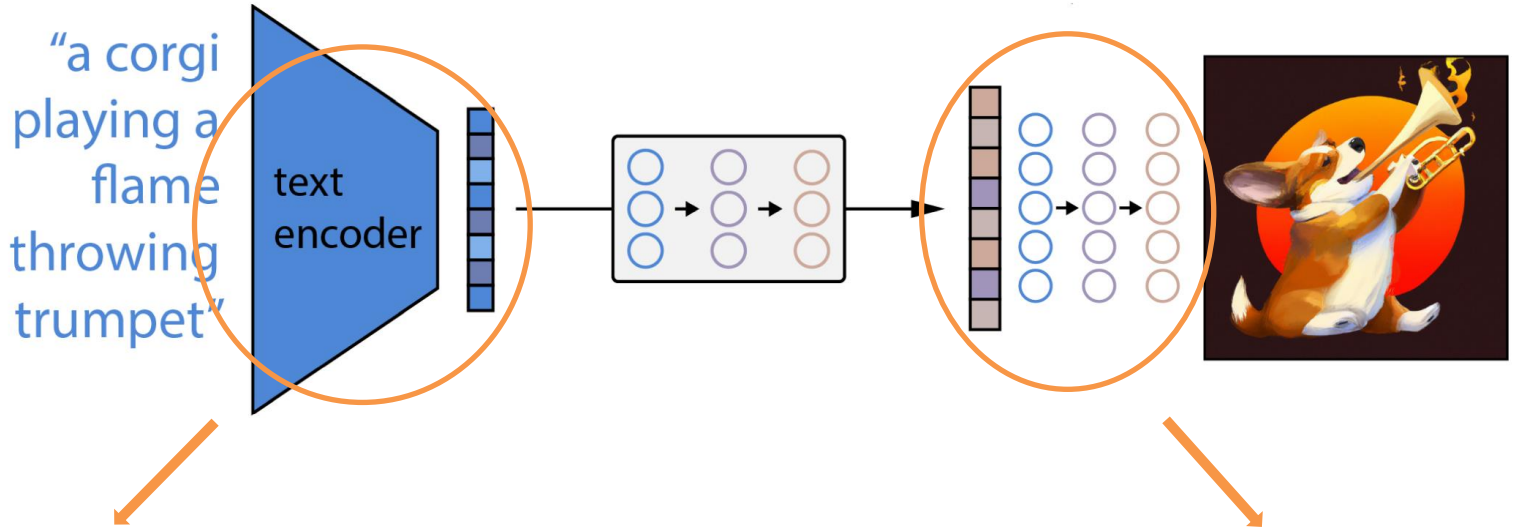
https://www.assemblyai.com/blog/how-dall-e-2-actually-works/

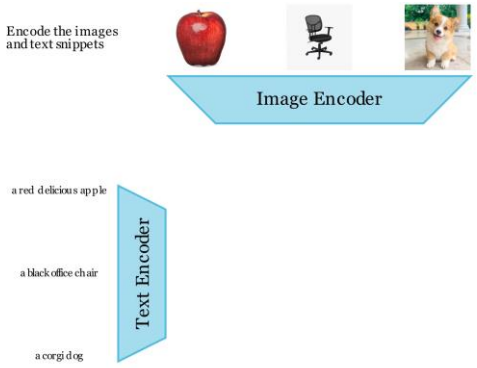
Para cerrar, veamos a muy alto nivel la arquitectura de DALL-E 2

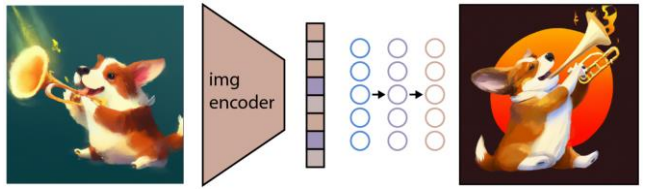


https://www.assemblyai.com/blog/how-dall-e-2-actually-works/

Para cerrar, veamos a muy alto nivel la arquitectura de DALL-E 2







https://www.assemblyai.com/blog/how-dall-e-2-actually-works/

Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ingeniería de Transporte y Logística

Sistemas Urbanos Inteligentes

Foundation Models

Hans Löbel
Dpto. Ingeniería de Transporte y Logística
Dpto. Ciencia de la Computación