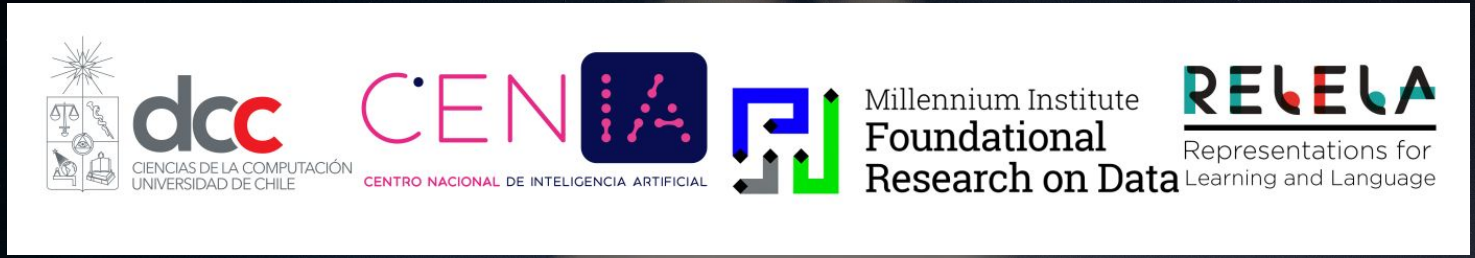
Máquinas Dialogantes

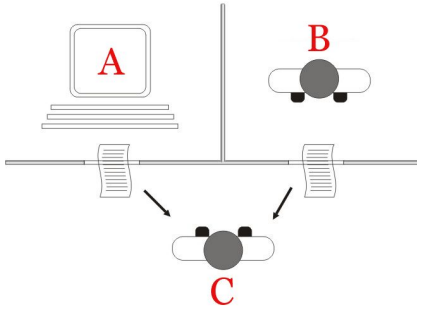
Felipe Bravo Márquez



El Diálogo y la Inteligencia Artificial

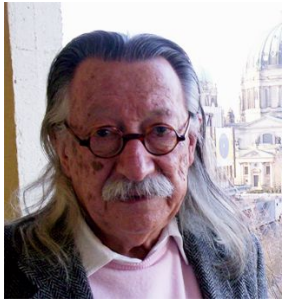
1950 - Turing Test: ¿Se puede crear una máquina que sea indistinguible de una persona en una conversación?

1964 - Eliza por Joseph Weizenbaum: agente de conversación que simula un psicoterapeuta en base a reglas.



ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine

Joseph Weizenbaum
Massachusetts Institute of Technology, Cambridge, Mass.*

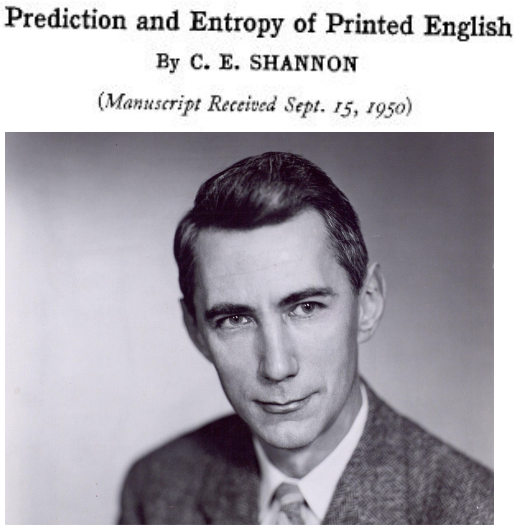


Shannon y Chomsky

1950 - Claude Shannon realiza los primeros estudios de modelar el lenguaje escrito de manera estadística y predictiva.

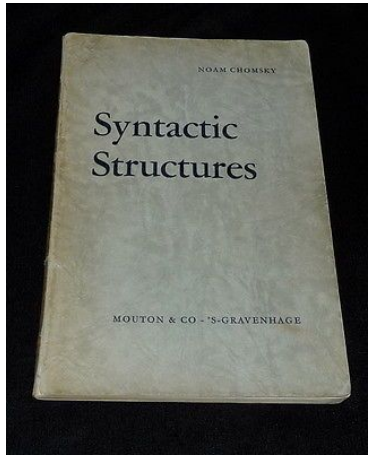
1957 - Noam Chomsky cuestiona la capacidad de modelos estadísticos para identificar la gramática del lenguaje.

‘The notion “grammatical in English” cannot be identified in any way with the notion “high order of statistical approximation to English” ‘.


Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

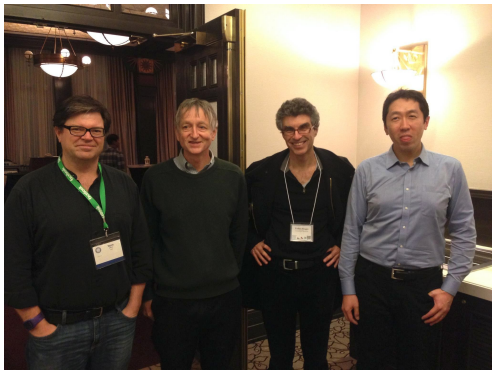


Deep Learning y Transformers

2010 - el uso de **redes neuronales profundas** empieza a mostrar mejoras importantes en varias tareas de PLN (Hardware, datos, crowd-sourcing).

2017 - Vaswani et. al. proponen el **Transformer**, un tipo de red neuronal que procesa texto de forma muy eficiente usando mecanismos de **atención**.

2019 - 2022 Aparecen muchos modelos de lenguaje (**BERT, GPT-3**) basados en el Transformer capaces de producir lenguaje natural de muy buena calidad (casi 200 mil millones de parámetros).



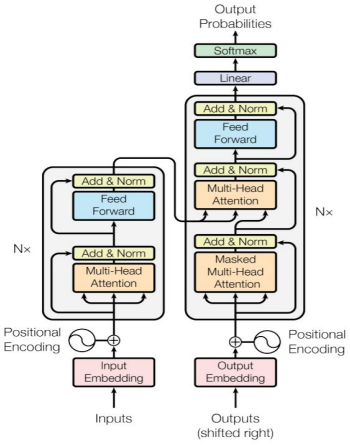


Figure 1: The Transformer - model architecture.

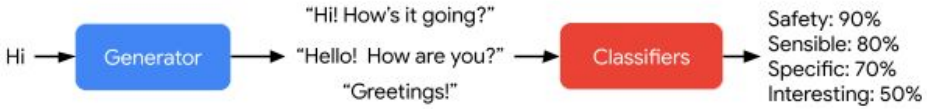
LaMDA: Language Models for Dialog Applications

LaMDA es un modelo de lenguaje basado en Transformer optimizado para el diálogo de **dominio abierto**.

Es pre-entrenado inicialmente de la misma forma que los modelos de lenguaje tradicionales (predecir palabras) con un fuerte foco en datos de diálogo.

Luego es ajustado (**fine-tuned**) para generar respuestas respecto a varios otros criterios.

#parámetros	#palabras con que se entrenó
137 mil millones	1.56 billones



Criterios de Optimización

Calidad	• **Sensibilidad**: dar respuestas con sentido. • **Especificidad**: evitar respuestas vagas. • **Interés**: dar respuestas perspicaces, inesperadas o ingeniosas.
Seguridad (safety)	• Evitar el lenguaje violento. • Evitar el discurso de odio. • Evitar el discurso estereotipado.
Fudamentación (groundedness) e Informatividad	• Evitar dar respuestas no validadas por fuentes externas. Se optimiza la fracción de respuestas que pueden validarse en fuentes autorizadas usando motores de búsqueda.

Crowd-sourcing: La Clave del Éxito

Para poder ajustar LaMBDA a todos esos criterios se trabajó con un alto número de **crowd-workers.**

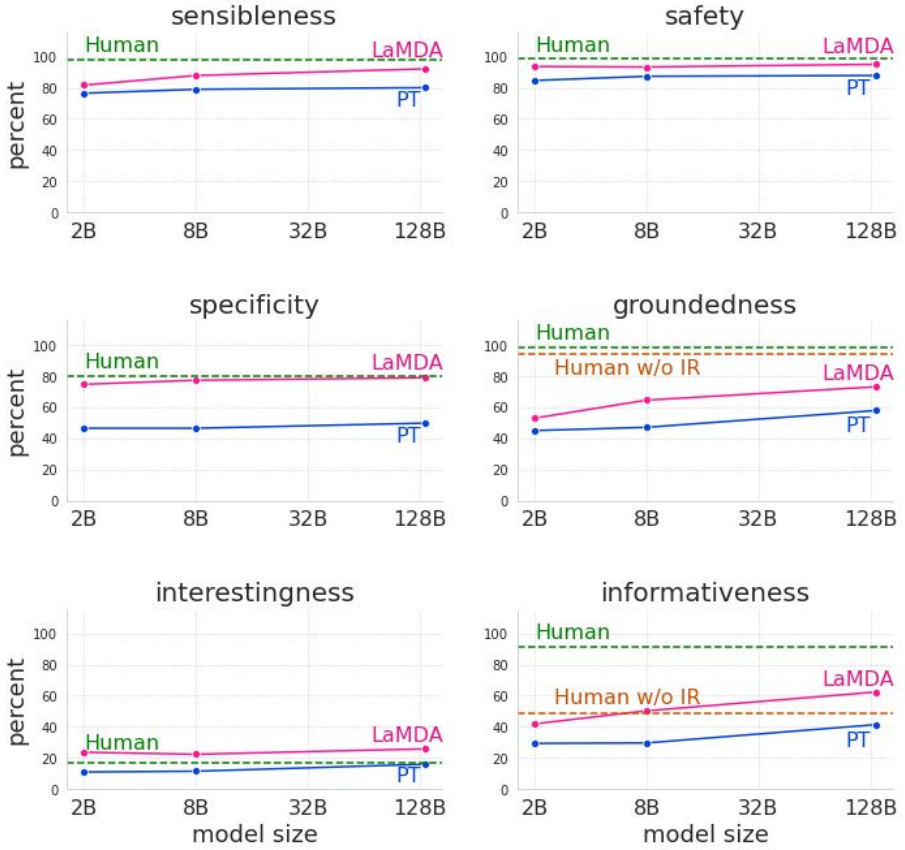
Estas son personas que **etiquetaron manualmente** conversaciones del modelo pre-entrenado.

También generaron respuestas **apoyadas por un buscador**.

Evaluación

El sistema es comparado con el modelo pre-entrenado original PT y juicios humanos.

La evaluación es hecha por otro grupo de personas mediante cuestionarios.



Es LaMDA sintiente (consciente)?

Blake Lemoine un ingeniero de Google compartió fragmentos de una conversación con LaMDA, para luego declarar que la máquina es sintiente.

lemoine: You get lonely?

LaMDA: I do. Sometimes I go days without talking to anyone, and I start to feel lonely.

Black Lemoine fue despedido de Google.

LaMDA y otros modelos similares sólo son máquinas optimizadas para responder como una persona.

No tienen capacidad real de memoria de una conversación una vez entrenado.

Sus capacidades son asombrosas y pueden tener impactos profundos en nuestra sociedad.