

Percepciones en IA 🤖

Un estudio sobre la percepción social de las capacidades lingüísticas y racionales de la Inteligencia Artificial General

Luca Mazzarello - Camila Migdal - Azul Noguera - Ignacio Pardo

Introducción

Meet 🤖 DAVID BLADE

El experimento lo protagoniza 🤖 DAVID BLADE, la AI/Robot contra la que se juzgan las palabras.

Test de Turing Minimalista

- A un juez se le presentan dos palabras A y B



- A partir de ellas, determina cuál la dijo una AI/Robot y cuál la dijo un humano.



- El objetivo del experimento a desarrollar es determinar la categoría más “robótica” entre dos.



Definimos una categoría como más “robótica” que la otra si frente al Test de Turing Minimalista es más probable que la haya dicho el Robot que el Humano.

Hipótesis

Dadas las 4 categorías seleccionadas para el estudio, hipotetizamos la relación entre todas ellas, determinando una de ellas más “robótica” que otra de la siguiente forma:

	>	
Tecnología		Afecto
Afecto		Lunfardo
Afecto		LDBP
LDBP		Lunfardo
Tecnología		Lunfardo
Tecnología		LDBP

Novedad

A diferencia del experimento original *A Minimal Turing Test* en el que basamos la investigación, nuestro experimento se enfoca en **distinguir** a 🤖 DAVID BLADE y no al humano o a ambos. De esta forma nos centramos en **clasificar** a las categorías de palabras como “Roboticas”.

Ademas, desarrollamos una encuesta capaz de **flexibilizarse** en cuanto a sus categorías de tal forma que pueda ser **replicable** sin estar sujeto a las categorías por las que optamos.

Categorías

Originales de “A Minimal Turing Test”

💖 **Afectos: (Affect)**
Empatía, Emoción, Sentimientos, Simpatía, Felicidad

🔥 **Life & Death, Bodily Functions and Profanity: (LDBP)**
Vivo, Dolor, Sexo, Caca

Agregadas a “DAVID BLADE”

📺 **Lunfardo: (Lunf)**
Birra, Boludo, Bondi, Che, Morfar

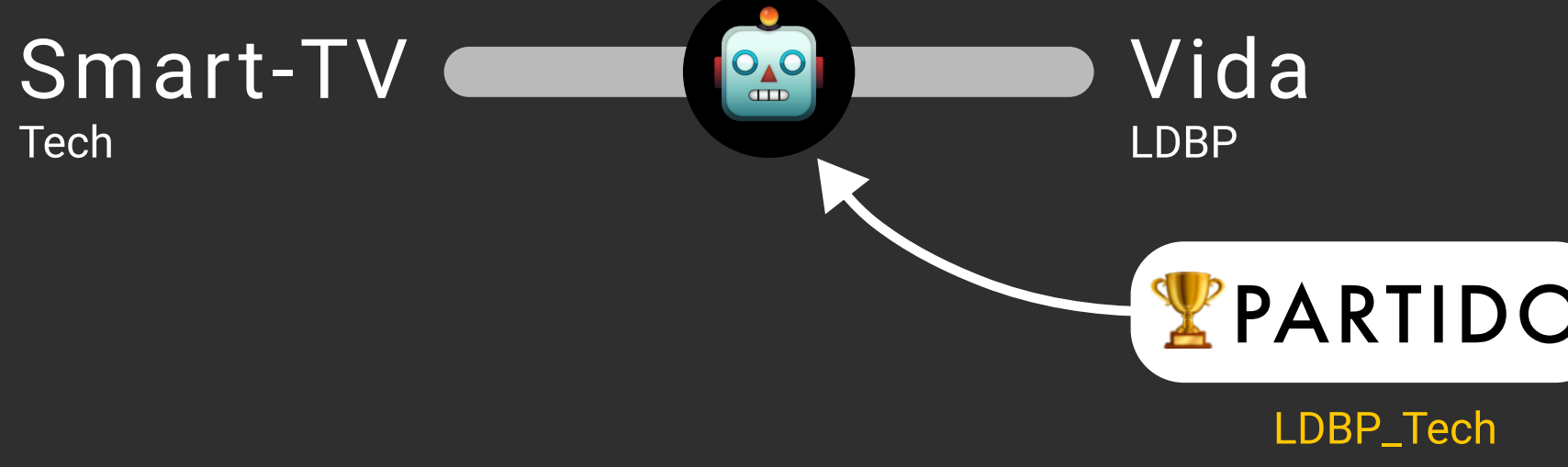
📱 **Tecnología: (Tech)**
Celular, Dron, Procesador, Robot, Smart-TV

Metodología

Interacción e Incentivo

Elegimos desarrollar un experimento interactivo no solo para atrapar al interesado, si no que tambien recompensar su tiempo con una escena de ciencia ficción a modo de entretenimiento. En ella, el participante toma el rol de juez en el cual determina el resultado de partidos entre 2 categorías cada una.

¿Qué palabra la dijo DAVID BLADE?



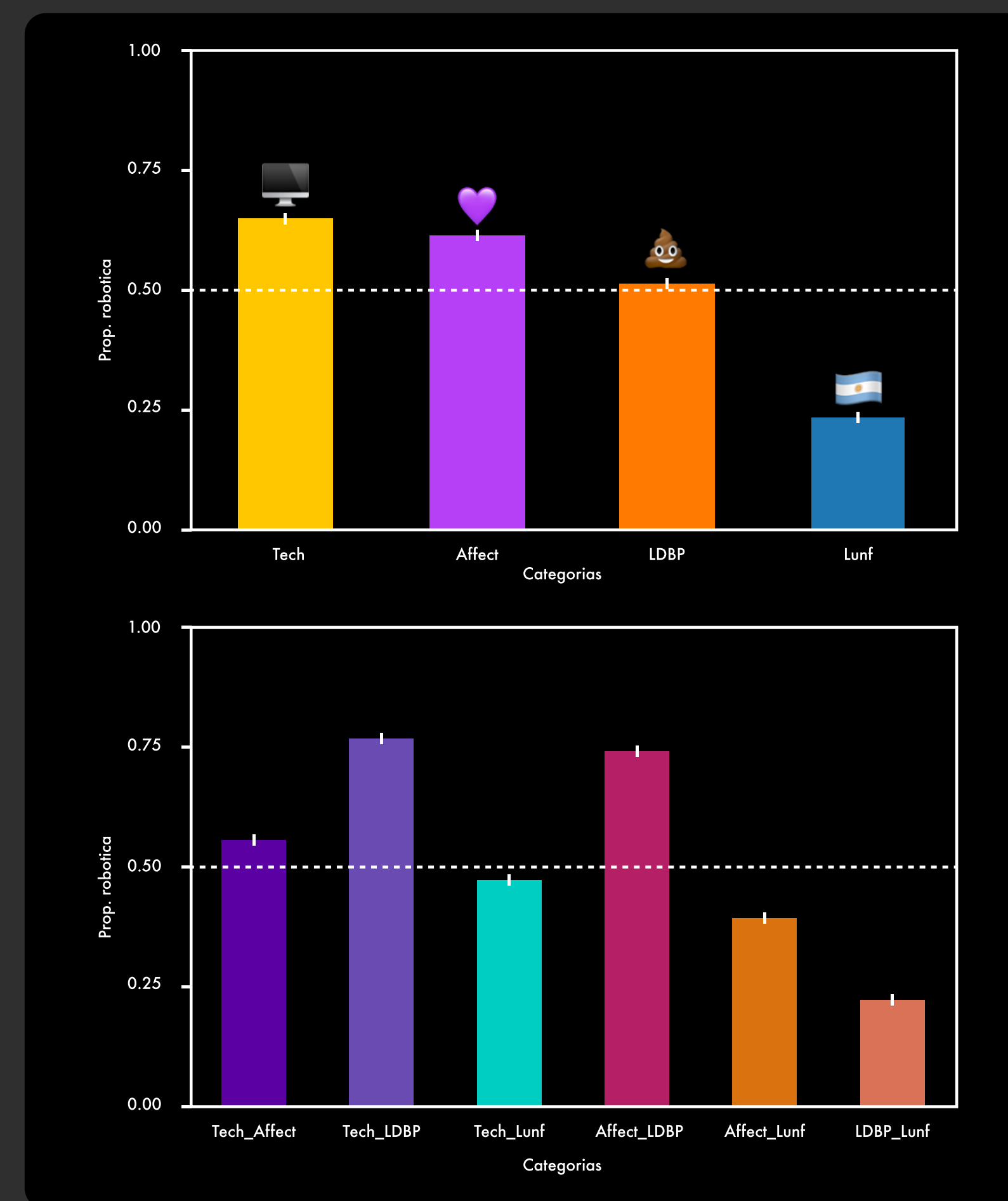
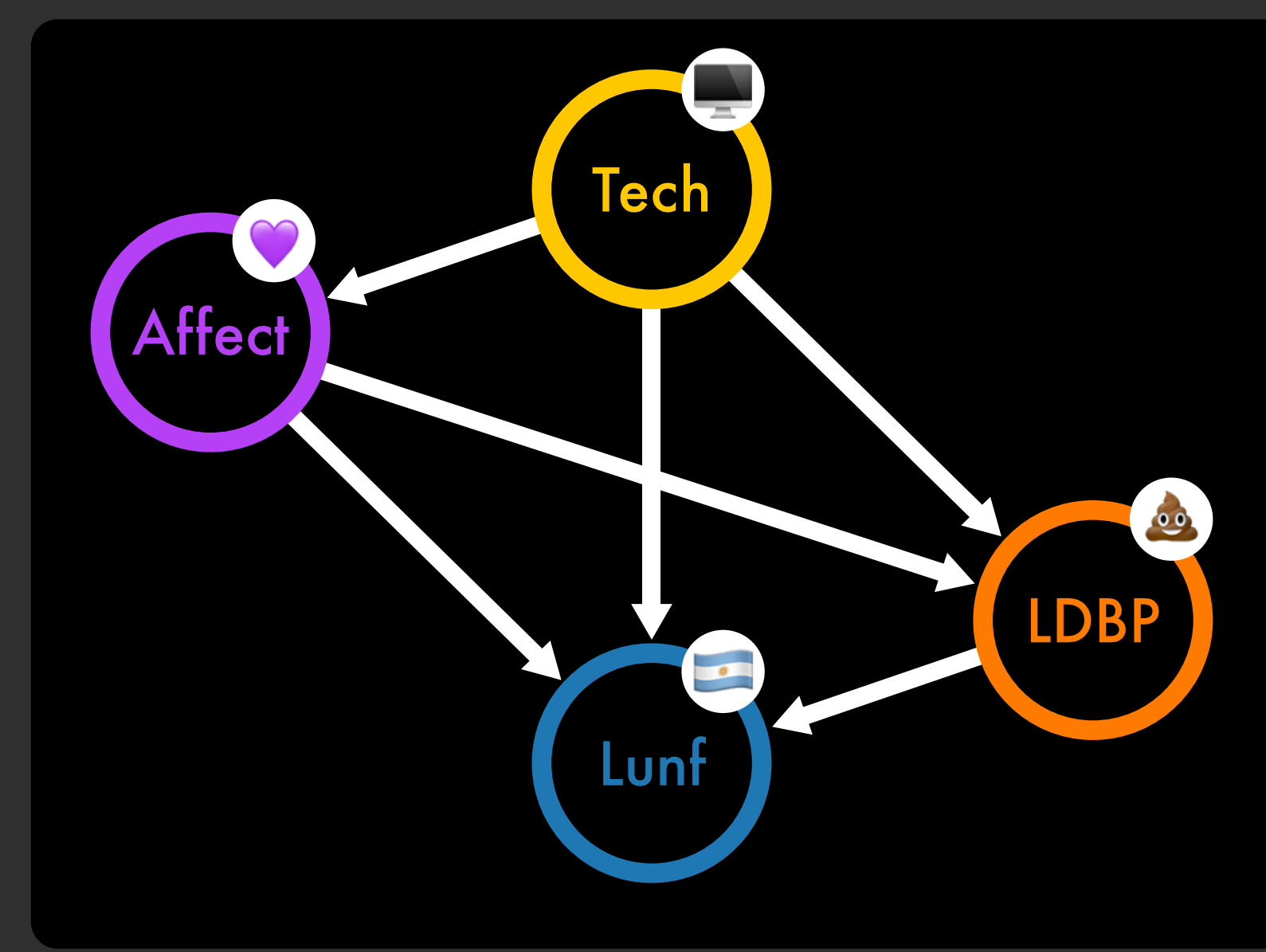
	Affect_Lunf	Affect_LDBP	LDBP_Lunf	LDBP_Tech	Lunf_Tech	Affect_Tech
Tech				211	427	260
Affect	539	378				228
LDBP		311	281	143		
Lunf	176		104		123	

La diferencia global de partidos ganados por una categoría contra a otra, nos dice de forma estadística cuan probable es que una palabra de una categoría la haya dicho 🤖 DAVID BLADE frente a otra.

Resultados

Relacion entre categorías

	Affect	LDBP	Lunf
Tech	53.278689	59.604520	77.636364
Lunf	75.384615	72.987013	
LDBP	54.862119		



Conclusión

AI computar

Respecto a la **hipótesis**, consideramos el experimento dio resultados semejantes a lo que anticipamos. Al ser un estudio **probabilístico**, la hipótesis del experimento no es confirmable, ya que de forma **inductiva** solo **nos podemos acercar a la hipótesis** a medida que más datos obtengamos.

	>	
Tecnología		Afecto
Afecto		Lunfardo
Afecto		LDBP
LDBP		Lunfardo
Tecnología		Lunfardo
Tecnología		LDBP



Bibliografía

McCoy, J. P., & Ullman, T. D. (2018). A Minimal Turing Test. *Journal of Experimental Social Psychology*, 79, 1–8. <https://doi.org/10.1016/j.jesp.2018.05.007>

UNIVERSIDAD TORCUATO DI TELLA