NOMBRE: Bernardita Rosas

SECCIÓN:

PUNTAJE:

Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ciencia de la Computación

IIC2613 — Inteligencia Artificial — 2' 2023

Tarea 2 – Respuesta Pregunta 1

Pregunta 1

Mucha gente argumenta que el podcast de Lex Fridman en YouTube es imperdible para todo quien se dedique a la inteligencia artificial. En mayo de este año, Lex publicó una entrevista a Stephen Wolfram, creador del answer engine Wolfram Alpha, que permite responder consultas lógicas a partir de información externa. Para contestar esta pregunta, primero ve la sección sobre la naturaleza de la verdad en el video (2:09:27 - 2:30:49).

1. En la entrevista, Lex propone que un modelo de lenguaje natural podría acercarse a tener una respuesta correcta a algunas preguntas subjetivas con suficiente información para su entrenamiento (por ejemplo, ¿es X persona buena?). Wolfram está en desacuerdo y plantea que la existencia de la verdad requiere de reglas o requisitos lógicos que la sustenten. ¿Estás de acuerdo con alguno de ellos? ¿Por qué? Explica tu punto de vista dentro de la discusión (puede ser el mismo que el de alguno de los oradores).

Estoy en bastante acuerdo con Wolfram, esto dado que al final todo depende de los datos con los que el modelo se alimenta y estos se encuentran inevitablemente sesgados. Como expresa Wolfram, se puede decir cierta "verdad" pero es realmente una posible verdad bajo reglas establecidas por los contextos de los dados. Lex incluso dice que hay "verdades" como no matar es malo, pero realmente si no existiera tal regla entonces se podría causar la extinción de los seres humanos y con ello ya nadie podría matar a otra persona, lo que no tiene gran sustento lógico. De todas maneras, existen y han existido culturas que realizan sacrificios humanos y, conforme al relativismo cultural, cabe preguntarse si realmente se tiene la verdad absoluta

sobre si matar es malo y si realmente podemos considerar nuestra verdad como absoluta al considerar nuestra cultura como "superior". Por ejemplo, ¿qué diferencia existencial existe entre matar a un animal ("bueno") y matar a un humano ("malo")? Me parece que siempre hay que cuestionar la verdad y creo profundamente que se puede encontrar "una verdad" más que la verdad absoluta.

2. ¿Crees que el concepto de verdad puede cambiar a lo largo de la historia? ¿Varía la verdad entre distintos contextos culturales? Da tres ejemplos que respalden tu respuesta.
Totalmente, creo que antes se creía que "la verdad" era un concepto más absoluto y de fé ciega mientras que ahora "la verdad" se basa mayoritariamente en la ciencia, lo lógico y lo demostrable. Por ejemplo, antiguamente se creía que todo giraba en torno a la tierra y hoy se "sabe" que esto no es así, sin embargo, personas como Galileo fueron perseguidas por no atingirse a algo que hoy se encuentra totalmente desvalidado. Incluso este cambio de paradigma demuestra de cierta forma lo difícil que establecer la verdad como algo que la IA podría descubrir. Nuevamente, creo que la verdad varía completamente entre culturas. En primer lugar, las religiones son un excelente ejemplo de cómo varía la verdad acorde a distintos contextos culturales. Mientras que para algunos existen múltiples dioses que nos crearon y gobiernan, para otros existe un Dios cuyo hijo vino al mundo o un profeta llamado Mahoma mientras que para otros no existe ningún Dios. Cada uno vive su propia verdad que es válida dentro de su contexto. En segundo lugar, respecto al aborto, es fácil comprender que es un tema muy sensible y genera opiniones muy diferenciadas. Mientras que para algunas personas corresponde a un asesinato, para otras corresponde a un derecho fundamental. Lo anterior viene dado por múltiples factores culturales y sociales que condicionan a las personas a inclinarse por una u otra respuesta. Nuevamente, no creo que nadie, ni mucho menos una inteligencia artificial, tenga autoridad absoluta para declarar una de las dos opiniones como absurda dado que cada persona tiene sus razones para opinar de cierta manera. En tercer lugar, los roles de género siguen siendo de gran controversia por su variación a través de distintas culturas. Por un lado, existen países en los que la mujer tiene las misma oportunidades

que un hombre mientras que hay otros en los que la situación es completamente opuesta y la mujer es obligada a vivir como ciudadana de segunda clase, totalmente dependiente de uno o más hombres a tal punto en el puede ser asesinada por intentar adquirir algunos de los derechos que tomamos por sentado en otras culturas. Definitivamente, me parece que la verdad cambia a través del tiempo y las culturas, al mismo tiempo que creo que nunca se podrá llegar a un consenso de verdad dadas las diferencias culturales.

3. Más adelante en la entrevista, Wolfram habla sobre el uso de LLMs como interfaces de comunicación y su uso para comunicar hechos (y ejemplifica con pedir un permiso para extraer peces), ¿cómo el uso de herramientas de lenguaje natural por uno o ambos extremos de la comunicación podrían significar la pérdida o cambios de información relevante e impactar a esta? Da tres ejemplos de escenarios con sus respectivas consecuencias.

El problema con el uso de tales herramientas viene dado que la inteligencia artificial se precupa más de que algo sea semánticamente posible que de que sea "verdad" según las reglas establecidas por la sociedad. Es necesario considerar que los LLMs no necesariamente distinguen entre leer ficción y un artículo científico, entonces se puede llegar a un montón de incongruencias y presentación de fantasía como real. En primer lugar, suele ocurrir que la inteligencia artificial simplemente inventa hechos o evidencias. El otro día, estaba hablando con una profesora y me contó que una alumna suya le pidió referencias a chat gpt para un tema relacionado con IA y se encontró con la sorpresa de que tales referencias no existían. Actualmente, al menos a mi conocimiento, la IA no crea documentos en internet, pero, ¿que pasaría si llegamos al punto en el que esto fuese plausible? En tal caso, perderíamos aún más la noción de realidad y no se podría confiar de nada de lo que se lee, lo que me parece extremadamente grave dado el volumen de "información" que la IA es capaz de producir. En segundo lugar, otro escenario que se me ocurre es el de una conversación con un LLM dedicada a diagnosticar a personas. El LLM no es capaz de leer correctamente que este se encuentra en un mal estado emocional (pierde esta información) por lo que no hace las preguntas correctas y le receta un remedio cuando la verdad este necesitaba

más un psiquiatra. La consecuencia de este escenario es que la persona no mejora su situación y compra un remedio que no necesita, despriorizando su bienestar. En tercer lugar, un escenario en el que se podría perder información es al usar traductores basados en LLM. Debido a los textos que alimentan a los LLMs, es poco probable que estos comprendan ciertos modismos cambiantes, parte esencial de la cultura de muchos lugares, incluyendo Chile, obviando el sentido en el que se está utilizando en un contexto dado. Bajo esta situación, podría ocurrir que se interprete una palabra como un insulto cuando la verdad corresponde correspondía a una muestra de confianza, generando un gigante malentendido y tal vez la pérdida de una oportunidad. Recapitulando, es necesario tener cuidado con los usos que les damos a los LLMs, ya que el mal uso de esto puede tener consecuencias extremadamente dañinas para el manejo de la información.