RHV, 2026, No 30, 45-59 eISSN 0719-4242
DOI: https://doi.org/10.22370/ rhv2026iss30.4679
Dossier Know-how / Articles

Saber cómo: testimonio experto, quasi-testimonios y *deepfakes*

Know-how: expert testimony, quasi-testimony and deepfakes

Felipe Álvarez[1]*

*Universidad de Chile
Fundación Centro Interdisciplinario en Ética, Política y Economía
felalvarez@uchile.cl
https://orcid.org/0000-0002-7153-2851

Resumen

Recientemente, la epistemología contemporánea ha puesto el foco en los *deepfakes* como un problema emergente. Sin embargo, no se ha abordado en detalle el *saber cómo* relativo a la distinción entre el testimonio experto y los *outputs* producidos por inteligencia artificial generativa. Este trabajo da cuenta de ese tipo de conocimiento señalando que consiste en ser capaz de mostrar, en la práctica, que uno distingue entre expertos virtualizados y pseudo-expertos digitales. Para dar cuenta de ello, en la primera sección se aborda la diferencia entre testimonio experto genuino y los quasi-testimonios producidos por IAG y manifestados en *deepfakes*; en la segunda sección se distingue entre el experto virtualizado y pseudo-experto digital que es IAG, señalando que el problema radica en suponer una equivalencia entre ambos; finalmente, se señala que el *saber cómo* relativo a distinguir entre ambos supone una sensibilidad digital que fundamenta que uno sea capaz de hacer acciones que muestren, efectivamente, que uno sabe cómo distinguir un *deepfake* de un experto real. Con ello, se busca contribuir a la discusión sobre el *saber cómo* relativo a *deepfakes*.

Palabras clave: *saber cómo*; testimonio; experto; IAG; *deepfakes*.

[1] Financiamiento: Agencia Nacional de Investigación y Desarrollo (ANID-PFCHA/Doctorado Nacional/Año 2022—21220627).

Received: 27/01/2026. Final version: 08/07/2026
eISSN 0719-4242 – © 2026 Instituto de Filosofía, Universidad de Valparaíso

Abstract

Recently, contemporary epistemology has focused on deepfakes as an emerging problem. However, the know-how concerning the distinction between expert's testimony and the outputs produced by a generative artificial intelligence has not been addressed in detail. This paper accounts for that kind of knowledge by pointing out that it consists in being able to show in practice that one distinguishes between virtualized experts and digital pseudo-experts. To account for this, the first section discusses the difference between genuine expert testimony and the quasi-testimonies produced by GAI and manifested in deepfake. The second section distinguishes between the virtualized expert and the pseudo-digital expert, pointing out that the problem lies in assuming an equivalence between the two of them. Finally, it is pointed out that know-how to distinguish between the two, presupposes a digital sensibility that demonstrates one's ability to take actions that effectively distinguish a deepfake from a real expert. Through this analysis, we aim to contribute to the discussion on the know-how related to deepfakes.

Keywords: know-how; testimony, expert, IAG; deepfakes.

1. Introducción

El auge de la inteligencia artificial generativa (IAG) ha dado lugar a diversos desafíos en cuanto a la adquisición y divulgación del conocimiento. Uno de ellos consiste en examinar los alcances y limitaciones de los denominados *deepfakes*, a saber, medios sintéticos (audio, video, entre otros) ultra realistas producidos por IAG. Algunos epistemólogos ya han cuestionado el papel que tendrá esta tecnología respecto de diversas fuentes epistémicas legítimas y cómo planeamos regularla finalmente (Fallis, 2020; Rini, 2020; Matthew y Kidd, 2023; Habgood-Coote, 2023). Sin embargo, no se ha puesto tanto el foco en caracterizar un *saber cómo* asociado a la identificación de los *deepfakes* por sobre los medios legítimos, de modo que una serie de problemas actuales como las estafas en la red potenciadas por los *deepfakes* o la desinformación científica derivada del uso (y abuso) de ellos, elevan la necesidad de caracterizar ese tipo de *saber cómo* y de examinar cuál es nuestra responsabilidad frente a esas tecnologías. En otras palabras, qué es lo que podemos hacer al respecto para evitar la confusión y diversos errores epistémicos al suponer una interacción humana cuando, en el fondo, solo ha habido interacción entre una persona y un *deepfake*.

Este artículo se hace cargo de ello haciendo una serie de distinciones conceptuales para dar cuenta de en qué consiste identificar un *deepfake* en la red en contraposición al testimonio legítimo de un experto virtualizado. Se pretende defender la idea de que el poder demostrar un *saber cómo* relativo a ello supone que uno comprende, en alguna medida, la distinción entre un *deepfake* y un experto real en la práctica haciendo acciones que den cuenta, aunque no

necesariamente de forma exhaustiva, de cuándo estamos frente a una persona o no. Para lograr dicho objetivo, este artículo se divide en las siguientes secciones: en primer lugar, en la sección 2, se analiza la diferencia entre los testimonios producidos por expertos con los supuestos testimonios producidos por IAG instanciados en *deepfakes*. Con ello se busca erradicar la posibilidad de equivalencia entre uno y el otro apelando a la normatividad socio-epistémica que rige el intercambio testimonial genuino, señalando que el tipo de confianza que tenemos hacia los expertos es radicalmente distinto al que tenemos con la IAG; luego, en la sección 3, se propone una distinción novedosa: la oposición entre expertos, expertos virtualizados, con pseudo-expertos virtualizados y pseudo-expertos digitales. Los tres primeros corresponderían a diversas formas de interacción entre personas, mientras que el último caracterizaría la interacción de un humano con un *deepfake* producido por IAG. Como se podrá entrever, el problema estará en suponer que se trata con un experto virtualizado cuando, de hecho, se trata con un pseudo-experto digital; finalmente, en la sección 4 se aborda el tipo de saber cómo relativo a distinguir entre el experto virtualizado y el pseudo-experto digital a partir del marco de la *interrogative capacity view* de Hagbood-Coote, indicando que para ese tipo de conocimiento es una de exigencia normativa el desarrollar de una especie de *sensibilidad digital* hacia los medios, ya sean legítimos o sintéticos, que uno se encuentra la red. Con ello se espera contribuir al debate epistemológico sobre el *saber cómo* relativo a los *deepfakes* como, a su vez, se procura hacer hincapié en la urgencia de actualizarnos para aprender a dominar las nuevas tecnologías emergentes derivadas de IAG.

2. Testimonio experto y quasi-testimonios

Confiamos en el testimonio experto en tanto que es prácticamente imposible justificar por cuenta propia cada una de nuestras creencias (Hardwig, 1985; 1991). Sin embargo, no es lo mismo confiar en una persona que en el *output* lingüístico de una IAG, inclusive si el enunciado en cuestión pareciera demostrar la misma pericia en el segundo caso que en el primero. En esta sección señalaremos las diferencias teóricas entre cada uno para luego, en la siguiente, señalar la clase de *saber cómo* que debería poseer un agente humano al interactuar con estas tecnologías. Con ellos mostraremos en qué consiste la problemática de suponer a un experto humano cuando simplemente nos encontramos con una IAG que lo simula, es decir, un pseudo-experto digital.

Una primera aproximación a esas diferencias la podemos encontrar en la oposición entre *confianza predictiva* y *confianza afectiva* de Paul Faulkner. De acuerdo con él, la primera consiste en lo siguiente:

> *Confianza predictiva*: "A confía en S para φ (en sentido predictivo) si, y sólo si, (1) A depende de S φ-endo y (2) A cuenta con S para φ (donde A cuenta con ello en el sentido de A predice que S hará φ)". (Faulkner, 2011, p. 145)[2]

[2] Todas las traducciones de artículos del inglés al español son de autoría propia.

De acuerdo con esto, la audiencia (el lego) espera que el hablante (el agente humano o IA) entregue un *output* verdadero en virtud de que es altamente probable que ese sea el caso. En otras palabras, se confía en el resultado, mas no en quién ha emitido ese resultado. La consideración de la confianza radicaría, por tanto, en examinar qué tan fiable es en un sentido veritista el agente en cuestión.

Esta clase de confianza viene bien a la hora de evaluar las capacidades epistémicas de cualquier clase de instrumento, al margen de qué tan complejo pudiera llegar a ser. Confío, por ejemplo, en un termómetro en tanto que es probable que el resultado que arroje sea verdadero. Bien podría haber algún caso Gettier que demuestre que he obtenido ese 'conocimiento' por suerte cuando el termómetro no ha funcionado bien, pero en situaciones estándar, no habría ningún problema en tener por fiable a ese instrumento y conocer a través de mi interacción con él siempre y cuando funcione como es debido. Lo mismo se puede decir, *mutatis mutandis*, de mi computadora o de una IAG como ChatGPT: confiamos en ellas solo en la medida de que son más conducentes al acierto que al error.

Si bien esta clase de confianza puede aplicarse también entre humanos, no se reduce nuestro modo de confiar en otros a una mera predicción de su comportamiento epistémico. Como menciona Faulkner, se puede confiar también de la siguiente manera, siendo este el estándar de la confianza entre personas:

> *Confianza afectiva*: "A confía en S para φ (en sentido afectivo) si, y sólo si, (1) A depende de S φ-endo; y (2) A cuenta con (1) para motivar a S para φ (donde A cuenta con ello en el sentido de que A espera de S que S sea motivado por la razón dada para φ)" (Faulkner, 2011, p. 146)

Así, cuando el lego confía en el experto, la dependencia epistémica del primero hacia el segundo debiera ser motivo suficiente para que este último testificase del mejor modo posible, a saber, de forma informativa, precisa y adecuada a las capacidades de su audiencia, El fundamento de esta clase de confianza radica en la normatividad social y en sus consecuencias. Si un médico me indica mal un remedio, puedo quejarme con la institución que le respalda. Inclusive, puedo demandar si el daño fuera demasiado grande. En cualquier caso, la palabra del experto opera como un garante epistémico en tanto que este no quiere defraudar a la audiencia por las posibles consecuencias que conlleva precisamente el hacerlo.

En el caso de las IAG's no existe este problema, pues no son responsables de sus 'conocimientos'. Al no tener estados mentales intencionales relativos a las proposiciones que emiten sus enunciados, son incapaces de entender a qué refieren con los términos involucrados y cómo la falta de precisión en ellos, o de plano el inventar información para complementar

una respuesta (como hace muchas veces ChatGPT y otros), podría generar un daño epistémico (y de otros tipos también) en su audiencia. De allí que la barra para medir estos instrumentos se limite, a pesar de su componente agencial, solo a sus resultados y no a cómo operan dentro de un entramado social complejo que exige determinadas disposiciones frente a la interacción con otros.

En una línea similar podemos encontrar la distinción entre creencias testimonialmente fundadas de las creencias tecnológicamente fundadas de Ori Freiman (2023). De acuerdo con él, si bien no se puede reducir a la IAG un mero instrumento (e.g. un termómetro) por su capacidad agentiva, eso no suscita que sea capaz de adherir subjetivamente a las normas que hacen del intercambio testimonial la clase de acto epistémico que, de hecho, es. Por lo tanto, la confianza que tenemos hacia los *outputs* lingüísticos de una IAG no es más que una estimación de la calidad de sus resultados, sin corresponder eso a lo que llamamos testimonio en realidad. Como bien menciona Freiman también, la distinción más bien debiera ser entre una 'IA fiable' más que una 'IA confiable', aunque el uso común (y no tan común) insista en la primera (Freiman, 2022).

Ahora bien, si lo anterior pone un punto de diferencia a nivel teórico entre el testimonio experto y el supuesto testimonio que podría dar la IAG, esto no nos libra de una dificultad práctica: ¿cómo discriminamos entonces a esos *outputs* lingüísticos que, sin ser testimonios, son capaces de engañar a la audiencia y aparecerse como testimonios genuinos? Esto será examinado en las siguientes secciones, no obstante, previo a ello tenemos que detenernos en una clasificación que sirve para ilustrar el punto: la distinción entre testimonio y quasi-testimonios.

Si bien el *output* lingüístico de la IAG no es un testimonio, sí que puede *pasar* por uno en la medida que, para la audiencia, en principio no habría diferencia significativa a nivel fenoménico entre un enunciado producido por una persona y por una IAG. Tal y como mencionan Miller y Freiman:

> Un *output* lingüístico de un instrumento o máquina constituye un quasi-testimonio en un contexto dado de uso si, y solo si, la máquina o instrumento ha sido diseñada y construida para producir este *output* de manera que recuerde suficientemente al testimonio fenoménicamente, y lo hace en conformidad con una norma epistémica de la cual es parasito, o suficientemente similar a lo que es, o debiera ser, una norma epistémica del testimonio en ese mismo contexto (2020, p. 429).

Esta adhesión a la norma epistémica que rige al testimonio fundamenta la confusión entre un testimonio real y un quasi-testimonio. En primer lugar, porque no podemos distinguir bien entre ambos; mientras que, en segundo lugar, inclusive si lo hacemos, suponemos que

podemos exigir en cualquier caso una retribución por el fallo al que nos podría inducir una IAG, lo cual no necesariamente es el caso. Consideremos el siguiente ejemplo para analizar esa cuestión:

> Caso 1: Juan y la IAG bancaria
>
> Juan requiere de asistencia bancaria dado que ha sufrido el robo de sus tarjetas de crédito y notó que fueron usadas para compras no reconocidas por él. Su banco, pionero en el uso de nuevas tecnologías para la atención al cliente, ha implementado una IAG programada para ayudar a quienes se encuentran en la situación de Juan, entre otras adversidades. Al llamar al banco, Juan es atendido por la IAG, la cual le da indicaciones erróneas a Juan. Juan sigue las indicaciones de la IAG, de modo que, luego, se encuentra en el siguiente problema: no logra dejar constancia efectiva del robo de modo que, aunque bloquearon sus tarjetas, el dinero perdido fue atribuido a un uso irresponsable de estas.

En el caso anterior uno podría decir que Juan no se ha dado cuenta, producto de su desesperación, que trató con una IAG en vez de con un ejecutivo bancario. Si bien puede parecer improbable, no es inusual que la población mayor sea incapaz de detectar si está tratando con esta clase de tecnologías, de modo que no podemos desestimar esta alternativa. Suponiendo que, entonces, se consumó el engaño, Juan podría culpar al banco por emplear una IAG que le ha conducido al error, diciendo que era necesario el testimonio experto de un ejecutivo de cuenta o del personal calificado correspondiente. El banco podría desentenderse del asunto por diversos motivos (la IAG estaba encargada a una empresa externa, por ejemplo), sin embargo, en estos casos parece más probable que el banco se haga cargo del error. Si bien no es responsable del *output* erróneo de la IAG, puede entregar una respuesta más satisfactoria a Juan.

Lo anterior parece ser un caso común, a saber, que un banco responda ante esos percances, pero ¿qué pasa cuando no se hace responsable una tercera entidad de la IAG y sus *outputs*? ¿Qué sucede cuando no sabemos diferenciar un quasi-testimonio del testimonio experto de una persona real? ¿Si no se encuentran fundamentados en la misma normatividad social que el testimonio, quién se hace responsable cuando la buena voluntad de una institución se queda corta, o cuando no hay regulaciones que obliguen a dar una respuesta satisfactoria al afectado? En la siguiente sección exploraremos más este punto a partir de dos nociones clave: la de pseudo-expertos digitales y la de *deepfakes*. Como se verá a continuación, ambas son formas en que la IAG se nos presenta cotidianamente en nuestro uso de la internet fenoménica (Smith, 2022), es decir, la internet del día a día. La ausencia de un *saber cómo* relativo a estas distinciones, propongo, es una de las causas predominantes de nuestros errores epistémicos al lidiar con la IAG.

3. Pseudo-expertos digitales y *deepfakes*

En nuestra cotidianidad, usamos la internet de diversas maneras para informarnos: leemos artículos científicos, leemos periódicos, consultamos foros, entre otros. Sin embargo, una forma de divulgación de información importante y, quizás, la más vigente en la actualidad, son las redes sociales (RRSS). El usuario común y corriente de internet se informa principalmente a través de ellas, pues allí confluyen distintos contenidos de su interés (lo que sea que quiera que fomente su *feed*, principalmente). Por ejemplo, bien puedo enterarme de la vida de mis amigos y sus novedades, a la vez que me entero de otros aspectos de la contingencia al desplazarme por la plataforma. Dado que la mayoría de los medios tradicionales cuentan con una versión digitalizada en RRSS, esto resulta ya natural y no solemos ver en ello alguna clase de problema.

Lo anterior es posible debido a que esos medios tradicionales, virtualizados, se han adaptado a las nuevas tecnologías y a las expectativas de los usuarios. No obstante, ha surgido un nuevo desafío para el testimonio experto en la red: los *deepfakes* y los posibles daños (epistémicos, morales, entre otros) que suscitan en nuestras sociedades. Esto se hace patente, sobre todo, considerando que la mayoría de *deepfakes* suele imitar a figuras de interés público (famosos, políticos, expertos en general, entre otros), por lo que confundirles puede suscitar percepciones erróneas de la ciudadanía hacia esas figuras y sus ideas, deformando con ello la opinión pública y la verdad respecto de sus posturas e ideas. Por consiguiente, para entender cómo nos puede llegar a afectar esto, debemos hacer una serie de distinciones previas a la caracterización de un *saber cómo* relativo a nuestra interacción con *deepfakes*. En esta sección distinguiremos, por tanto, entre el testimonio experto virtualizado y lo que llamo 'pseudo-testimonio digital', el cual es producido por IAG e instanciado alternativamente por medio de *deepfakes*.

El testimonio experto se puede virtualizar porque la función según la cual 'X' cuenta como 'experto' en 'la sociedad globalizada' no está atada materialmente a su portador, sino que existe más bien como una representación que es aceptada por el resto de la comunidad (Brey, 2003; Searle, 1995; 2010). Las credenciales que acreditan esa función, es decir, aquel dispositivo epistémico que nos permite identificar al experto (Smith et al, 2020), también se pueden reproducir en la red, por lo que podemos señalar al portador como lo que realmente es a nivel social-institucional sin que haya diferencia alguna entre la credencial física y la credencial digital. Así, si me topo con una conferencia (o el extracto de una conferencia) de un físico teórico en Instagram como, supongamos, Michio Kaku, sabré que lo que dice es un testimonio experto sobre el tema expuesto y que, por tanto, es información fiable. En mi posición de ignorante, podría llegar a aprender sobre teoría de cuerdas escuchando ese testimonio de forma asincrónica inclusive. Por lo tanto, el testimonio experto sobrevive a la virtualización

en una primera instancia y constituye una buena práctica para la divulgación del conocimiento.
Otra forma de virtualización del testimonio experto consiste en la posibilidad de generar una interacción sincrónica con él. Siguiendo un ejemplo parecido al anterior, uno podría imaginarse conectándose a una transmisión en vivo (un *live,* como se suele denominar) para interactuar con el experto en cuestión. Esto, sobre todo post pandemia del Covid 19, se estandarizó como una buena alternativa a toda la logística que implica el producir y asistir presencialmente a una conferencia como tal. En este segundo escenario, la virtualización es, inclusive, mejor que las instancias presenciales para la divulgación de conocimiento, pues las limitaciones físicas que nos alejan y no nos permiten interactuar, dejan de estar presentes, por lo que hasta podemos suponer que es una alternativa preferible en muchos contextos.

Ahora bien, tal cual como se instancian virtualmente los expertos, lo hacen los pseudo-expertos. Un divulgador de pseudociencia puede colarse y hacerse pasar como un experto en la red. De este modo, nos encontramos con dos clases de pseudo-expertos: por una parte, el pseudo-experto malintencionado que imita con conciencia al experto para obtener un rédito personal, a saber, dinero u otra clase de bien no epistémico en la interacción con otros; mientras que, por otra parte, tenemos al pseudo-experto ignorante que, producto de un sesgo tipo Dunning-Kruger (Kruger y Dunning, 1999), se tiene a sí mismo como alguien con determinada pericia aunque no haya sido aprobado su supuesto conocimiento por una institución certificadora competente. En cualquier caso, la audiencia cae en ellos porque estas formas patológicas de interacción se nutren de la norma epistémica del testimonio, es decir, yendo todo normal, la audiencia tiende a creer y tener razones epistémicas para creer *en alguien* presencialmente o en la red.

Todo lo anterior se ve complementado con un nuevo problema: la suposición de pericia hacia entidades agentivas no humanas. Si hacerle caso a un pseudo-experto virtualizado es un problema epistémico ya importante (sobre todo considerando las consecuencias no epistémicas derivadas de la ignorancia y el error), ahora debemos sumarle que el auge de la IAG ha suscitado que las personas atribuyan pericia a diversos programas de procesamiento de lenguaje natural. Analicemos el siguiente caso para examinar el punto:

> Caso 2: Pedro y el uso de IAG para aprender
>
> Pedro es un estudiante de primer año en una prestigiosa universidad chilena. *Ad portas* de terminar el primer semestre, se cuestiona si es necesario seguir asistiendo a clases porque, según él, ChatGPT explica mejor los ejercicios de álgebra que su profesor. Al consultar con su docente, él le responde que, a diferencia de ChatGPT, él es un experto certificado, pero Pedro sugiere que, si ChatGPT logra dar una mejor explicación matemática, entonces es igual o más experto que su profesor.

En este caso, se considera como experto a una IAG. Como vimos en el apartado anterior, esto es un error desde el comienzo debido a que ningún programa puede, en estricto rigor, testificar. Sin embargo, que recuerde fenoménicamente el quasi-testimonio de la IA a un testimonio real (que el discurso de ChatGPT sobre álgebra sea *similar* al testimonio experto del profesor universitario), suscita que luego se puedan generar casos como recién expuesto. Esto, lejos de ser descartado como una mera actitud irracional por parte de Pedro, debe analizarse con mesura, pues constituye una postura común actualmente frente al desarrollo de la tecnología. Por esa razón, es importante distinguir a los *outputs* producidos por IAG que simulan pericia de los meros discursos de los charlatanes que simulan expertos en entornos virtualizados, de modo que propongo llamar los primeros bajo el término de pseudo-expertos digitales, es decir, una clase de entidad agentiva cuya existencia es meramente digital y simula ser un experto por la clase de enunciados que es capaz de producir. Esto, como ya se podrá notar, se opone al pseudo-experto virtualizado en la medida que el primero es simplemente un pseudo-experto instanciado en la red. Sin embargo, esta nueva definición nos permite caracterizar esta nueva clase de fenómenos que parasitan la norma epistémica del testimonio y que son cada vez más comunes.

En el caso 2, no obstante, si bien hay confusión respecto de lo que significa ser un experto para Pedro (pues atribuye erróneamente pericia a una IAG cuando, en estricto rigor, no pueden tener pericia), no se ha confundido a la entidad agentiva no humana que es la IA con una entidad agentiva humana como lo podría ser un experto virtualizado. En ese sentido, solo habría que corregir la consideración de Pedro respecto de la supuesta pericia de la IAG. Sin embargo, ¿qué sucede cuando no se puede distinguir a la IAG de una entidad agentiva humana genuina con pericia acreditada o, en otras palabras, cuando se confunde al pseudo-experto digital con el experto virtualizado? Para abordar ese punto, debemos hablar acerca de los *deepfakes*.

Los *deepfakes* son imágenes, audios o videos que, creados con IAG y otras herramientas, simulan de forma realista el comportamiento de personas reales. Así, uno puede generar un video en el que Donald Trump, actual presidente de los Estados Unidos, declare que subirá los aranceles de importación a China un 300%. Dado que el comportamiento de Trump es poco predecible, bien que alguien podría, si el tono de voz es el adecuado y las imágenes suficientemente convincentes, suponer que el genuino Donald Trump ha dicho eso y tomar por testimonio un pseudo-testimonio producido por una IAG.

En el ámbito del testimonio experto y la pericia, esto también puede suceder. Supongamos una variación del caso 2 para ilustrar el punto:

Caso 2.1: Pedro y el uso de *deepfakes* para aprender

> Pedro, *ad portas* de terminar el primer semestre, se cuestiona si es necesario seguir asistiendo a clases porque, según él, ha encontrado un profesor en YouTube que explica mejor los ejercicios de algebra que su profesor. Al consultar con su docente, él le responde que es un experto certificado, a lo que Pedro responde que el profesor de YouTube también lo es, puesto que ha visto una imagen de su certificado de grado de doctor en matemáticas, similar al que podría mostrarle su docente universitario. Sin saberlo, no obstante, Pedro ha estado estudiando con un *deepfake* que simula ser un profesor universitario cuyo discurso es producido, al igual que el resto de los elementos en ese material audiovisual, por una IAG.

Como se podrá ver, ahora el problema no es que Pedro suponga pericia en un programa, sino que no sabe que el programa es, de hecho, un programa. Si bien es cierto que hay buenos profesores explicando materias universitarias en plataformas como YouTube u otras, podría ser el caso que Pedro justo diese con un *deepfake* que simula ser un docente dando clases en internet (supongamos, un *deepfake* que alguien ha puesto a dar clases para lucrar con las *views*). Esa IAG, como se nota en el caso, también es capaz de generar pseudo-credenciales para engañar a su audiencia en caso de ser requerido; y dada nuestra poca disposición a verificar la información y a considerar como una pieza de evidencia suficiente meras imágenes u otros materiales (siempre y cuando calcen con nuestro *background* de creencias) con una sola mirada, bien que podríamos caer en diversos *casos Gettier peligrosos*, a saber, casos en los cuales, en circunstancias estándar sin elementos anómalos azarosos, tendríamos en efecto una creencia verdadera justificada sin tener *conocimiento* (Hetherington, 2016, p. 9). De ese modo, parece inevitable caer en el error de confiar en un pseudo-experto digital como se confía en un experto virtualizado, suponiendo allí la existencia del segundo y no la del primero.

Lo anterior parece ser un ejemplo rebuscado, pero no está tan lejos de la realidad siendo que los *deepfakes* han adquirido cada vez más presencia en la red. Sin embargo, no se ha hecho hincapié en el *saber cómo* relativo a la interacción con esa tecnología, sobre todo para evitar la gettierización derivada de esta clase de casos. Dado que, comúnmente, se defiende una incompatibilidad entre la caracterización del saber cómo en relación con la posibilidad de la ser *gettierizado*, en lo que procede, nos dedicaremos a caracterizar ese *saber cómo* sobre la base de las distinciones previamente hechas para mostrar cómo los usuarios pueden evitar caer presos del engaño al aprender a detectar medios sintéticos como producciones hechas con IAG, específicamente con el caso de los *deepfakes*.

4. *Saber cómo* distinguir al pseudo-experto digital del experto virtualizado

De acuerdo con Habgood-Coote (2019), el *saber cómo* se puede caracterizar como "una habilidad para activar conocimiento detallado para responder a una pregunta en una serie de situaciones prácticas contextualmente dadas en las que uno activa ese conocimiento haciendo

un tipo de actividad relevante para ello" (p. 92). En otras palabras, como lo resume Morales Carbonell:

> En un contexto determinado c, S sabe cómo hacer x si y solo si, dadas las situaciones prácticas $F1…Fn$ asociadas al contexto, S tiene la capacidad de activar su conocimiento de una respuesta detallada a todas (o la mayoría) las preguntas de la forma "¿Cómo hacer x en Fi?", al intentar hacer x. (2025, p. 108)

Si aplicamos esto a nuestro tópico, a saber, indicar el tipo de conocimiento que alguien ha de tener para distinguir cotidianamente entre el pseudo-experto digitalizado y sus quasi-testimonios del experto virtualizado y su testimonio genuino, podemos tener una caracterización adecuada de ello. Analicemos este nuevo escenario para dar cuenta de ello:

Caso 3: Manuel y sus padres víctimas de estafa

Manuel se ha enterado que sus padres donaron una suma importante de sus ahorros a un supuesto médico; en efecto, han visto en Facebook un video en el que Luis Peralta, médico cirujano, ha declarado necesitar dinero para patentar su novedosa vacuna contra el cáncer de mamas. Ante esto, Manuel sospecha y les comenta: "A propósito de los *deepfakes* que están tan de moda, ¿cómo saben siquiera que ese médico existe?". El padre de Manuel responde "tranquilo, es real, yo lo sé", por lo que procede a mostrarle el video a su hijo. Manuel, preocupado, concluye que sus padres no saben distinguir cuando están frente a un *deepfake* y que, efectivamente, han caído presos de una estafa.

En el escenario podemos observar que, si bien los padres de Manuel pueden responder afirmativamente a la pregunta acerca de si saben o no distinguir un *deepfake*, carecen de un genuino *saber cómo* al respecto. Eso se debe a que no se ha activado un conocimiento relativo a generar esa distinción por medio de la acción de mostrar el video al hijo. El saber cómo distinguir *deepfakes* de testimonios reales implica más que solo afirmar que uno puede hacerlo. Implica, por ejemplo, hacer acciones que, si quisiéramos poner en términos de proposiciones, sirvan como una respuesta en el contexto en el que manifiesta esa clase de preguntas. Manuel, que sí posee ese *saber cómo*, podría mostrarles a sus padres que en el video hay diversos elementos audiovisuales no coinciden con lo que esperaríamos de un video real (e.g. la representación de los dedos, la prosodia de la voz, entre otros), sospecha que surge por el contexto en el cual se ha obtenido acceso a ese material (a saber, las RRSS). Podría, a su vez, señalar que no es común que un médico solicite dinero por sus perfiles personales, ni menos para ese tipo de causas, googleando diversas campañas de *crowdfunding* para causas benéficas o de interés ciudadano lideradas por profesionales de la salud para mostrar formas legítimas de llevar a cabo esa recaudación de fondos. Inclusive, más alentado por su sospecha, podría hacer *fact cheking* para contrastar ese video con fuentes fiables que den cuenta de la posible estafa o, en su defecto, acudir a testimonios de otros afectados en RRSS, entre otras tantas

alternativas. En cualquier caso, esas demostraciones activarían en él un conocimiento del cual sus padres carecen y que respondería a la pregunta sobre cómo distinguir entre el pseudo-experto digital y el experto virtualizado haciendo, precisamente, esas demostraciones.

En este caso podemos notar también que los padres de Manuel representan una postura común respecto de estas tecnologías, la cual es, a su vez, uno de los mayores problemas sobre estos temas: la gente cree que, de hecho, es capaz de responder a la pregunta sin mayores problemas solo por el hecho de usar cotidianamente internet. Sin embargo, esto está lejos de la realidad y podemos suponer una carencia de *humildad intelectual* respecto de estas nuevas tecnologías, i.e, una falta en la capacidad de evaluar el estatuto epistémico de las propias creencias (Church y Samuelson, 2017). Esto lo podemos apreciar en la investigación de Köbis et al (2021), quienes han señalado que las personas tienen dos clases de sesgos respecto de sus propias habilidades para estimar si se encuentran frente a un *deepfake* o no: en primer lugar, tienden a identificar como auténticos aquellos videos que no lo son a pesar de haber sido informados previamente de que se distribuirán de forma equitativa los videos reales con aquellos que sí son *deepfakes*; mientras que, en segundo lugar, tienden a subestimar su propia capacidad para distinguir *deepfakes* de videos reales, es decir, caen en el efecto Dunning-Kruger que anteriormente atribuimos a los pseudo-expertos ignorantes. En otras palabras, nos creemos más versados de lo que realmente somos respecto de estas tecnologías.

Si bien esto nos muestra que las personas tienden a sobrestimar sus capacidades y que, por tanto, ignoran su falta de *saber cómo* al respecto, esto no debe ser malinterpretado: al igual que uno no se vuelve guitarrista solo por tomar una vez en la vida una guitarra (y tocarla, probablemente, mal), uno tampoco desarrolla habilidades para distinguir esas tecnologías interactuando esporádicamente con ellas. En el caso de Manuel, podemos suponer que, por una cuestión generacional, ha de estar más acostumbrado que sus padres a ver estos casos e identificarlos. En términos de Taylor Matthews, habría adquirido una mejor *sensibilidad digital*, a saber, un mejor ajuste en su de radar emocional relativo a la interacción con material audiovisual en la red, incluyendo videos legítimos como *deepfakes* (2022). Si esto lo llevamos al caso de sus padres, simplemente quedaría pendiente ajustar esa sensibilidad en pos de desarrollar la habilidad de discriminar *deepfakes* tal y como lo hace su hijo, es decir, que logren identificar, con un mayor grado de éxito, al pseudo-experto digital del experto virtualizado a través del cultivo constante de un sano escepticismo hacia los medios digitales. Es, por tanto, una cuestión por practicar.

Lo anterior nos lleva al siguiente punto: hay una responsabilidad epistémica en la audiencia respecto de qué tan susceptibles son de ser engañados (o permanecer en el engaño) una vez que han tomado por verdadero un video hecho con IAG, sobre todo considerando los tiempos que corren. Hoy en día uno no puede simplemente argüir ignorancia respecto de las tecnologías que están permeando nuestras sociedades para eludir la culpa de no estar a la altura del *saber cómo* relativo a ello. El caso 2.1 y el caso 3 dan cuenta de escenarios en los que

se esperaría de los protagonistas que sepan discernir entre el trato con un experto virtualizado y un pseudo-experto digital, pero no logran hacerlo. Dado que gran parte de nuestra formación y adquisición de información se está realizando en ambientes digitales, se vuelve fundamental que sepamos discriminar entre material audiovisual el que sí hay un testimonio experto genuino y, por tanto, garantía respecto del valor epistémico del contenido que se transmite a diferencia de los quasi-testimonios producidos por la IAG. Aunque los últimos no necesariamente son conducentes al error, no participan de la normatividad socio-epistémica del mismo modo que un testimonio experto genuino, por lo que bien podríamos caer en ellos sin tener que haya más responsables que nosotros mismos al confiar en una herramienta cuyo valor, como se ha visto, no es equiparable al de un humano (Álvarez y Espinosa, 2024). Dado que este tipo de saber es acumulativo y progresivo, será nuestra tarea el ajustar la propia sensibilidad digital para lograr responder a la pregunta sobre cómo distinguir expertos virtualizados de pseudo-expertos digitales sin caer en los sesgos cognitivos que, precisamente, nos hacen suponernos como poseedores de ese saber cuando, de hecho, no lo poseemos.

5. Conclusiones

A lo largo de este trabajo nos hemos abocado a caracterizar el *saber cómo* que ha de tener alguien para evitar caer en *deepfakes*. Partimos señalando en la sección 2 la imposibilidad de obtener testimonio experto genuino de la IAG en tanto que ella no suscribe, directamente, a la normatividad socioepistémica como lo hacen los expertos, de modo que confiamos en sus *outputs* solo en un sentido predictivo. Así, a pesar de haber allí una entidad agentiva, ello no es suficiente para determinar que los quasi-testimonio producidos por ella sean equivalentes a los testimonios genuinos producidos por expertos que sí participan a esa normatividad.
En la sección 3 defendimos que el rol social del experto se puede virtualizar, incluyendo con ellos las credenciales que certifican su pericia, por lo que el testimonio que derive de él es prácticamente idéntico y, en determinados escenarios, preferible al intercambio testimonial fuera de ambientes digitales. A pesar de esto, indicamos que, los ambientes virtuales son lugares propensos para la propagación de deepfakes, los cuales pueden operar como pseudo-expertos digitales para engañar a la audiencia. A diferencia de un pseudo-experto virtualizado, en el caso de los *deepfakes*, no estaríamos interactuando con una persona, sino con el *output* de una IAG que ha sido diseñada para inducirnos al error, tal y como lo señalamos en las variantes del caso 2.

Finalmente, en la sección 4, caracterizamos el *saber cómo* relativo a la distinción entre pseudo-expertos digitales y expertos virtualizados a partir de la *interrogative capacity view* de Habgood-Coote. Esto implicaría manifestar la activación de un determinado conocimiento relativo a la resolución de la pregunta sobre si estoy o no frente a una IAG, o el producto de una IAG, en lo que respecta a la adquisición de conocimiento en la red. Asimismo, señalamos que, a pesar de que caigamos en sesgos y podamos erróneamente atribuir la existencia de un experto virtualizado donde hay un pseudo-experto digital, es exigible de nosotros, por los tiempos

que corren y el desarrollo continuo de estas tecnologías, el desarrollar una determinada sensibilidad respecto de los *deepfakes*, la cual debemos ir afinando para evitar caer en el error de suponer interacción humana mutua (con todo lo que normativamente eso suscita) cuando, de hecho, no la hay.

En síntesis, se ha argumentado que el *saber cómo* distinguir *deepfakes* surge en la medida que uno es capaz de hacer acciones que demuestren que uno comprende la diferencia entre interactuar con expertos virtualizados y con pseudo-expertos digitales en la red; y que la falta epistémica que se deriva de la ausencia de ese tipo de *saber-cómo* radica en la suposición errónea de atribuir humanidad a una IAG siendo que esta no opera bajo el mismo marco de referencia que los expertos virtualizados. Esa cuestión sería, en parte, nuestra responsabilidad.

Como se podrá entrever, en este artículo parece recaer todo el peso en el usuario. Sin embargo, solo se ha querido recalcar el *saber cómo* referido a uno y su interacción con estas tecnologías digitales. Ello no niega que distintas instituciones deban fomentar un buen uso de la IAG y procurar, en la medida de lo posible, el cumplimiento de buenas normas asociadas a la producción y divulgación de *deepfakes*. En una sociedad cohesiva y preocupada de las consecuencias de la IAG y los *deepfakes*, de sus alcances y limitaciones éticas y epistemológicas, ambas cosas debieran ir de la mano.

Referencias bibliográficas

Álvarez, F y Espinosa, R. (2024). "The value of testimonial based beliefs in the face of AI generated quasi-testimony". *Aufklärung 11* (especial): 25-38. https://doi.org/10.18012/arf.v11iEspecial.70023

Brey, P. (2003). "The Social Ontology of Virtual Environments". En Koepsell, D. y Moss, L. (eds.). John Searle's Ideas About Social Reality: Extensions, Criticisms and Reconstructions. Oxford: Blackwell. https://doi.org/10.1111/1536-7150.t01-1-00011

Church, I. and Samuelson, P. (2017). Intellectual Humility: An Introduction to the Philosophy and Science, London: Bloomsbury.

Kruger, J., and Dunning, D. (1999). "Unskilled and unaware of it: how difficulties in recognizing ones own incompetence lead to inflated selfassessments". *Journal of Personality and Social Psychology 77*(6), 1121-1134. https://doi.org/10.1037/0022-3514.77.6.1121

Faulkner, P. (2011). *Knowledge on trust*. Oxford: Oxford University Press.

Fallis, D. (2020) "The epistemic threat of deepfakes", *Philosophy and Technology, 34*(4): 623–643. https://doi.org/10.1007/s13347-020-00419-2

Freiman, O. (2022). Making sense of the conceptual nonsense 'trustworthy AI'. *AI Ethics* 3, 1351–1360. https://doi.org/10.1007/s43681-022-00241-w
Freiman, O. (2023). Analysis of Beliefs Acquired from a Conversational AI: Instruments-based Beliefs, Testimony-based Beliefs, and Technology-based Beliefs. *Episteme*, *21*(3), 1031–1047. https://doi.org/10.1017/epi.2023.12
Habgoode-Coote, J. (2019). "Knowledge-How, Abilities, and Questions". *Australian Journal of Philosophy 97*(1): 86-104. https://doi.org/10.1080/00048402.2018.1434550
Habgoode-Coote, J. (2023). "Deepfakes and the epistemic apocalypse", *Synthese, 201*(103): 1–23. https://doi.org/10.1007/s11229-023-04097-3
Hardwig, J, 1985, "Epistemic dependence". *The Journal of Philosophy 82*: 335-349. https://doi.org/10.2307/2026523
Hardwig, J, 1991, "The role of trust in knowledge". *The Journal of Philosophy 88* (12): 693-708. https://doi.org/10.2307/2027007
Hetherington, S. (2016). *Knowledge and the Gettier Problem*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316569870
Köbis et al. (2021). "Fooled twice: People cannot detect deepfakes but think they can". *iScience 24*(11). 103364. https://doi.org/10.1016/j.isci.2021.103364
Matthews, T. (2022). "Deepfakes, Intellectual Cynics, and the Cultivation of Digital Sensibility". *Royal Institute of Philosophy Supplement 92*: 67-85. https://doi.org/10.1017/S1358246122000224
Matthews, T y Kidd, I. (2023). "The ethics and epistemology of deepfakes". En Fox and Saunders (eds.) *The Routledge Handbook of Philosophy and Media Ethics*. London: Routledge. https://doi.org/10.4324/9781003134749
Miller, B. y Freiman, O. (2020). Trust and distributed epistemic labor. En: Simon, J. (Ed.) *The Routledge handbook on trust and philosophy*. New York: Routledge. https://doi.org/10.4324/9781315542294
Morales Carbonell, F. (2025). *¿Cómo? Una introducción a la epistemología del saber cómo*. Montbretia.
Rini, R. (2020) "Deepfakes and the epistemic backstop", *Philosopher's Imprint 20*: 24. http://hdl.handle.net/2027/spo.3521354.0020.024
Searle, J. (1995). *The Construction of Social Reality*. London: Penguin Books.
Searle, J. (2010). *Making the Social World: The structure of Human Civilizations*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780195396171.001.0001
Smith, B., Guliana Loddo, O. y Lorini, G. (2020). On Credentials. *Journal of Social Ontology 6* (1): 47-67. https://doi.org/10.1515/jso-2019-0034
Smith, J. (2022). *The Internet Is Not What You Think It Is: A History, a Philosophy, a Warning*. Princeton: Princeton University Press. https://doi.org/10.2307/j.ctv1z3hmcj