

## Inteligencia Artificial en la creación de falsos documentales en las aulas de arte: desafiando la desinformación a través de narrativas pseudo-históricas y deepfakes

Felipe Aristimuño 

Universidade de Lisboa, Faculdade de Belas-Artes, Centro de Investigação e de Estudos em Belas-Artes (CIEBA), Largo da Academia Nacional de Belas-Artes, 1249-058 Lisboa, Portugal, [felipe-aristimuno@campus.ul.pt](mailto:felipe-aristimuno@campus.ul.pt)

Breve bio autor: Doctor en Bellas Artes por la Universidad de Lisboa (FBAUL 2020), con especialización en Educación Inclusiva (FMHUL 2022). Investigador en multimedia y educación artística, afiliado al Centro de Investigación y Estudios en Bellas Artes (CIEBA).

How to cite: Aristimuño, F. (2024). Inteligencia Artificial en la creación de falsos documentales en las aulas de arte: desafiando la desinformación a través de narrativas pseudo-históricas y deepfakes. En *libro de actas: EX±ACTO. VI Congreso Internacional de investigación en artes visuales ANIAV 2024. Valencia, 3-5 julio 2024*. <https://doi.org/10.4995/ANIAV2024.2024.17698>

---

### Resumen

*El surgimiento de deepfakes ha generado inquietud respecto al robo de identidad y la desinformación, cuestionando los límites de la inteligencia artificial (IA). Sin embargo, hay un creciente interés en su uso para ofrecer perspectivas críticas a través del arte. En este contexto, presentamos el proyecto "Oeiras Voltada ao Mar" de la Escuela Secundaria Quinta do Marquês en Oeiras, Portugal, desarrollado entre enero y mayo de 2023. Este proyecto involucró a estudiantes de tercer año de secundaria en la exploración del potencial de la IA en la representación del patrimonio local y las narrativas históricas, centrándose en el uso de ChatGPT y D-ID para generar deepfakes. Los estudiantes colaboraron introduciendo indicaciones en ChatGPT para crear relatos pseudo-históricos sobre los faros de Oeiras, que luego fueron presentados por celebridades a través de deepfake. Los videos resultantes se exhibieron en el evento "Somos Todos Proyecto" en junio. En este ensayo, reflexionamos acerca de los resultados de esta experiencia, destacando los eventos de alucinaciones de IA presenciados por los estudiantes, además de la protección proporcionada por D-ID contra el uso no autorizado de imágenes de celebridades internacionales en deepfake.*

**Palabras clave:** Inteligencia Artificial; GPT; Deepfake; Falso documental; Desinformación.

---

### Abstract

*The emergence of deepfakes has sparked concerns surrounding identity theft and misinformation, pushing the boundaries of artificial intelligence (AI) into question. Nevertheless, there is a growing interest in leveraging AI to offer critical perspectives through art. With this in mind, we introduce the "Oeiras Voltada ao Mar" project, conducted at Quinta do Marquês Secondary School in Oeiras, Portugal, between January and May 2023. This initiative engaged third-year high school students in exploring AI's potential in depicting local heritage and historical narratives, specifically focusing on the use of ChatGPT and D-ID for generating deepfakes. Students actively participated by providing prompts to ChatGPT to craft pseudo-historical narratives centered around the lighthouses of Oeiras, which were subsequently portrayed by celebrities through deepfake technology. The resulting videos were showcased at the "Somos Todos Projeto" event in June. Within this essay, we reflect on the results of this experience, highlighting the occurrences of AI hallucinations witnessed*

*by students, as well as the protection provided by D-ID against unauthorized use of international celebrity images in deepfakes.*

**Keywords:** Artificial Intelligence; GPT; Deepfake; Mocumentary; Fake news; Misinformation.

## INTRODUCCIÓN

Los *deepfakes* han sido identificados como una nueva forma de desinformación, planteando un riesgo para la privacidad y la seguridad. Sin embargo, existe una gran incertidumbre acerca de cómo esta tecnología puede ser utilizada con propósitos beneficiosos, como estimular la creatividad en el ámbito artístico y en la educación mediática. En respuesta, nos sumergimos en el proyecto “Oeiras Voltada ao Mar”, realizado por la Escuela Secundaria Quinta do Marquês en Oeiras, Portugal. Esta iniciativa tenía como objetivo envolver a estudiantes de tercer año de secundaria para explorar formas de representar el patrimonio local y las diversas posibilidades de narrativas históricas a través de la imagen.

En este artículo, analizaremos el papel de las clases de artes visuales en el proyecto, donde se empleó el modelo de lenguaje *Generative Pre-trained Transformer* ([www.chatgpt.com](http://www.chatgpt.com)) junto con la plataforma D-ID ([www.d-id.com](http://www.d-id.com)) para crear videos *deepfake*. Los estudiantes dieron indicaciones a ChatGPT para generar relatos pseudo-históricos sobre los faros de Oeiras, utilizando luego *deepfake* para presentarlos a través de una celebridad, explorando así las implicaciones de los *deepfakes* en la proliferación de noticias falsas.

Como referencia para la planificación, consultamos dos documentos: el informe ético de la Comisión Europea sobre el uso de la Inteligencia Artificial (IA) en la educación (CE, 2022), que nos proporcionó pautas éticas para la intervención humana en la IA y la transparencia; y el Referencial para la Educación en Medios (MEC, 2014), que en el contexto portugués aborda la educación mediática como un campo transdisciplinario, capacitando a los ciudadanos para analizar críticamente la comunicación contemporánea.

Mediante un diseño metodológico enfocado en la investigación en el aula (Hopkins, 2014), orientado hacia el currículo, la enseñanza y las perspectivas de los estudiantes, y utilizando un enfoque *Art Thinking* (Acaso & Megías, 2018), nos proponemos descubrir cómo utilizar eficazmente la IA para crear obras de arte mediático en el aula de artes, además de explorar sus limitaciones éticas y su potencial reflexivo en el contexto educativo.

## DESARROLLO

### 1. Motivos para la integración de *deepfakes* en el aula de artes

Para explorar la integración de la IA en el aula de artes, es fundamental entender el concepto de *deepfake*. Según Laura Payne (2024), los *deepfakes* son medios sintéticos generados por IA que representan algo inexistente en la realidad. El término, que combina “*deep*” (profundo) y “*fake*” (falso), se popularizó en 2017 en un subreddit de Reddit que utilizaba el intercambio de rostros para insertar imágenes de celebridades en otros videos (Fig. 1).

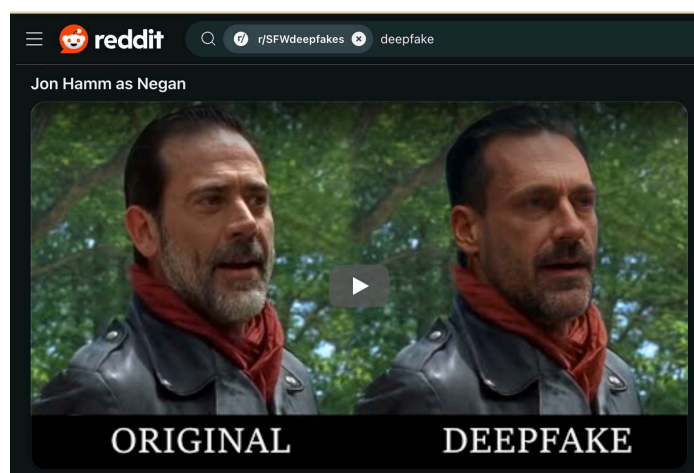


Fig. 1: Deepfakes en Reddit. Fuente: <https://www.reddit.com/r/SFWdeepfakes/>

Como han destacado académicos como Weikmann (2022) y Seow et al. (2022), los deepfakes constituyen una forma de desinformación que amenaza la privacidad, la seguridad e incluso las bases de la democracia. Estos estudios subrayan los daños causados por su uso malicioso y la necesidad urgente de implementar medidas de detección eficaces.

Como herramienta creativa, los artistas de medios han recurrido a los *deepfakes* para explorar la relación entre humanos y máquinas, realidad y fantasía, privacidad y vigilancia. Entre las obras destacadas se encuentran “This is the Future” de Hito Steyerl (2019), “Delusional Mandala” de Lu Yang (2017), y “The Flowers” de Dries Depoorter (2022). Además, los memes de *deepfake* (Fig. 2) desempeñan una función de activismo visual, aumentando la conciencia sobre la propaganda digital y la industria de la influencia.



Fig. 2: ejemplos de memes de deepfakes: (a) El rostro de Jim Carrey transferido al cuerpo de Alison Brie, (b) Mr. Bean es Charlize Theron en una versión Deepfake del comercial de J'adore, (c) Jim Carrey en lugar de Jack Nicholson en *El Resplandor*, y (d) Tom Cruise reemplaza a Robert Downey Jr. en *Iron Man*. Fuente: Guarnera, Guidice & Battiato (2020)

Como educadores, hemos notado un creciente interés en la IA en nuestras escuelas, impulsado por la popularización de la IA generativa. El arte siempre ha creado ilusiones de la realidad, desde la perspectiva renacentista y las esculturas de dioses griegos hasta los vitrales góticos. Creemos que los *deepfakes*, de manera similar, pueden tener influencias tanto positivas como negativas: son una amenaza para la identidad, pero también pueden servir como crítica social, algo fundamental en el aula de arte contemporáneo. Nos corresponde

reflexionar sobre cómo utilizar eficazmente la IA en la creación de arte mediático, considerando sus limitaciones éticas y potenciales.

## **2. La IA Generativa en el Proyecto “Oeiras Voltada ao Mar”**

Para enfrentar estos desafíos, participamos en el proyecto “Oeiras Voltada ao Mar” entre enero y mayo de 2023. Dirigido por el Club del Patrimonio de la escuela y respaldado por la asociación cultural “Espaço e Memória”, este proyecto involucró a nuestros alumnos en asignaturas como Lengua Portuguesa, Inglés, Historia y Artes Visuales. El objetivo era explorar diversas formas de representar los históricos faros de Oeiras, utilizando desde poesía y fotografía hasta maquetas.

Desde las clases de arte, y basándonos en el modelo ArtThinking (Acaso & Megías, 2019), identificamos la IA generativa como un “pensamiento divergente” (Ibid.), ya que es algo nuevo e incierto en términos educativos, lo que genera dudas, miedo y curiosidad en nuestros alumnos y en nosotros mismos. Hemos contrapuesto esta divergencia con la convergencia: si la IA puede generar historias e imágenes, ¿por qué no puede servirnos para representar nuestro patrimonio cultural de manera crítica y en el contexto de los problemas actuales del mundo? Así, propusimos a los estudiantes que colaboraran introduciendo indicaciones en ChatGPT para crear relatos pseudo-históricos. Luego, que seleccionaran a una celebridad para presentar sus videos usando *deepfake* (Fig. 3), explorando así las implicaciones de la IA generativa tanto como medio de producción creativa como de generación de *fakenews*.



**Fig. 3: Captura de pantalla de video - Maria Leal, conocida cantante de música popular portuguesa, presentada en una recreación deepfake junto al faro de São Julião da Barra, Oeiras. Fuente: Trabajos de alumnos en clase (2023)**

Para nuestra investigación en clase, hemos seguido la “Guía del Profesor para la Investigación en el Aula” de Hopkins (2014), adoptando una mirada crítica hacia nuestra práctica para reflexionar sobre nuestras acciones y la calidad de los resultados. Siguiendo el enfoque de investigación-acción de Hopkins, nuestra práctica se dividió en tres etapas: 1) analizamos la necesidad de abordar la creatividad mediante IA en nuestro contexto educativo, 2) implementamos acciones para ello produciendo contenidos en *deepfake* y 3) reflexionamos sobre los resultados para proponer nuevas ideas.

### 3. Actividades en clase: Un resumen de las actividades realizadas

El proyecto se desarrolló en ocho sesiones de noventa minutos cada una, con la participación de dos clases de tercer año de secundaria, totalizando 45 alumnos, de edades entre 12 y 14 años, incluyendo cuatro alumnos con necesidades educativas especiales.

En la primera sesión, se aclararon conceptos sobre IA generativa mediante videos de artistas destacados (Steyerl, 2019; Yang, 2017; Deepporter, 2022) y debates, utilizando también material del curso en línea gratuito “Alfabetización mediática en la era de los *deepfakes*” del MIT (Harrel et al., 2021). Se exploró el concepto de *deepfake* y su evolución, promoviendo la reflexión con preguntas sobre sus impactos futuros, sus posibilidades creativas en las artes y cómo distinguirlos de imágenes o videos reales. Luego, se enfocaron en el trabajo práctico, utilizando ChatGPT como asistente para redactar guiones de videos, con énfasis en la ética y el respeto.

En los equipos de la biblioteca escolar (Fig. 4), se propuso a los estudiantes la creación de *prompts* para elaborar guiones destinados a videos breves, con leyendas inventadas sobre los faros de Oeiras en un estilo juvenil y al modo de un *youtuber*.



Fig. 4: Trabajo práctico en la biblioteca. Fuente: Elaboración propia.

Un ejemplo de *prompt* para introducir en ChatGPT sería:

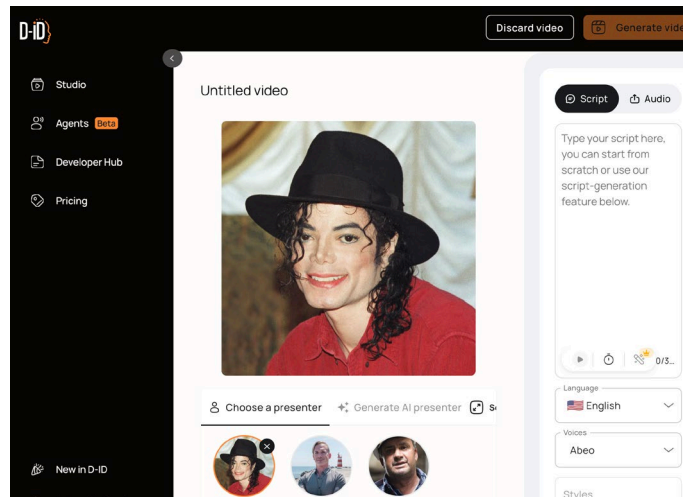
“Me gustaría que escribieras un guion corto para un video de dos minutos, con un tono juvenil, que cuente una leyenda sobre (a elegir por el grupo de alumnos) en el faro (elegido por el grupo de alumnos) de Oeiras, ocurrida poco después de su inauguración. ¿Podrías incluir algo de información histórica sobre el faro y su ubicación? También sería interesante mencionar cómo el faro es visitado en la actualidad y el impacto que tiene la leyenda en la comunidad.” (Elaboración propia)

Con los guiones generados por ChatGPT, los grupos discutieron posibles cambios en los textos y verificaron la exactitud de los datos históricos. Luego, cada grupo eligió una celebridad fallecida o una figura pública portuguesa no protegida por derechos de imagen para su video *deepfake*. El objetivo no era suplantar identidades, sino emplear la cultura del remix (Lessig, 2008) con fines humorísticos y de crítica social.

Es importante destacar que el informe ético de la Comisión Europea sobre el uso de la IA en la educación subraya la protección de las identidades y los datos personales de todos los involucrados en el proceso educativo (CE,

2022). Por eso, el concepto de *Fair Use* (uso justo) defendido por Lessig (2009) es fundamental para reflexionar sobre la apropiación de imágenes de manera justa, sin perjudicar a nadie.

Utilizamos la aplicación D-ID para crear los videos *deepfake* a partir de fotografías (Fig. 5). Aunque es más parecido a un *shallowfake*<sup>1</sup>, fue adecuado para nuestros propósitos de remix, proporcionando una experiencia similar al *deepfake* con un toque humorístico.



**Fig. 5: La interface de D-ID.** Fuente: Trabajos de alumnos en D-ID

Después de crear los videos, los estudiantes finalizaron sus ediciones en la clase de Tecnologías de la Información y la Comunicación (TIC), añadiendo sonidos y superponiendo imágenes. Compartieron los videos en Microsoft Teams para que sus compañeros los vieran y comentaran. Finalmente, reunimos una colección de 21 videos, de aproximadamente dos minutos cada uno, y los presentamos en la Feria de Proyectos el 1 de junio de 2023 en la escuela.

#### **4. Reflexión acerca de los resultados**

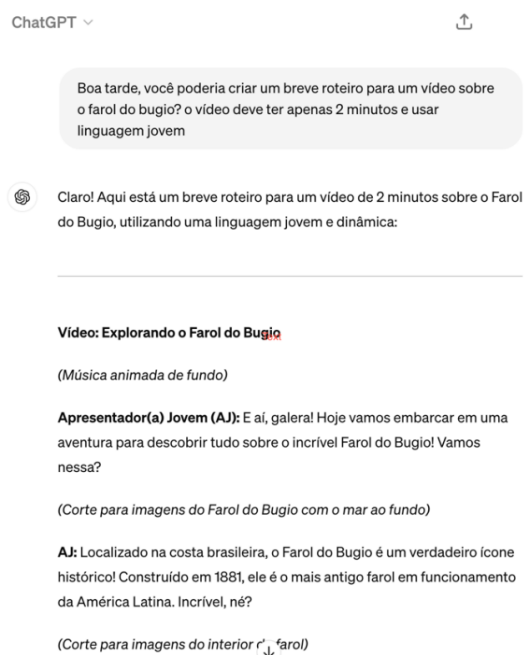
En el debate realizado, aunque el término "*deepfake*" no era novedoso, la mayoría pensaba que su creación requería recursos costosos. Al ver los videos seleccionados, las opiniones se dividieron sobre si el uso de *deepfake* era realmente perjudicial, destacando que las creaciones no eran creíbles por la voz robotizada y el movimiento imperfecto de los labios. Se explicó que la tecnología aún está evolucionando.

Surgieron preocupaciones sobre el uso de *deepfake* para el bullying, dañar la imagen de personas famosas y manipular elecciones. En cuanto al arte, no hubo respuestas afirmativas, incluso con ejemplos de artistas que utilizan *deepfake*. Sobre su detección, los alumnos destacaron nuevamente la baja calidad de los montajes. Se enfatizó la necesidad de investigar fuentes creíbles y desconfiar de imágenes aisladas en redes sociales.

Sobre la utilización de ChatGPT, muchos alumnos ya lo habían experimentado en casa, motivados en gran medida por los debates previos. Surgieron comentarios sobre el temor de otros profesores al uso de la IA para falsificar composiciones. Dado que ChatGPT no recomienda su uso a menores de 18 años sin supervisión, se sugirió que los alumnos menores usaran cuentas de profesores. Varios guiones presentaron "alucinaciones de la IA", como

<sup>1</sup> Un *shallowfake* es una forma de manipulación de medios que no utiliza tecnologías avanzadas de inteligencia artificial o aprendizaje automático (...). Este método, al ser menos intensivo tecnológicamente, es más accesible para la persona promedio, lo que permite un uso generalizado. Sin embargo, la naturaleza manual de los shallowfakes a menudo resulta en alteraciones menos convincentes en comparación con los *deepfakes*. (Deepfakenow, 2023)

el caso en que ChatGPT ubicó incorrectamente el Faro del Bugio en Brasil (Fig. 6). Estos errores destacaron la importancia de verificar los hechos.



**Fig. 6: Captura de pantalla de ChatGPT – interacción de alumnos con la IA.**  
Fuente: Trabajos de alumnos en clase (2023)

A pesar de esta lección, una pareja no verificó su guion y realizó un video con información falsa sobre el faro. Decidimos no corregir el error para que los estudiantes reflexionaran sobre la fiabilidad de la IA. Los alumnos compartieron sus guiones en Microsoft Teams bajo el tema “Mockumentaries”, presentando al todo 21 guiones sobre faros. Las leyendas incluían fantasmas en el Faro del Bugio, una sirena en el Faro de São Vicente y una criatura alienígena en el Faro de Gibalta.

Sobre la creación de los deepfakes, las figuras seleccionadas incluyeron al youtuber SirKazzio, la modelo Margarida Corceiro, la cantante María Leal, el youtuber Wuant, la Gioconda de Da Vinci, Michael Jackson, el político ultraconservador André Ventura y un hombre que se ha convertido en meme por inyectarse aceite en los brazos (Fig. 7). Gracias a la sencillez de uso de la aplicación, todas las parejas completaron la tarea con facilidad.



**Fig. 7: Capturas de pantalla de videos – Recreaciones de personalidades en deepfake junto a faros de Oeiras.**  
Fuente: Trabajos de alumnos en clase

Después de ajustes y refinamientos, los videos fueron presentados en la Feria de Proyectos de la escuela, generando gran interés entre todas las clases. Aunque los deepfakes no eran completamente realistas, captaron la atención de los visitantes, quienes permanecieron hasta el final de las presentaciones.

## CONCLUSIONES

A juzgar por el compromiso de todos los participantes —alumnos, profesores y audiencia— pudimos concluir que hay un interés creciente en nuestro entorno educativo por las tecnologías de IA, tanto en la generación de texto como de imágenes y sonidos. Existe una comprensión de la importancia de estas tecnologías en términos de futuro. Sin embargo, notamos que todavía no hay una comprensión clara acerca de sus peligros ni de las potencialidades en términos artísticos y pedagógicos. Nuestras limitaciones en relación con la apropiación de imágenes de otras personas tampoco nos parecieron suficientemente claras. De hecho, esta es una discusión que no ha comenzado con la IA, sino con la cultura del Remix desde hace casi veinte años (Lessig, 2009).

En este sentido, consideramos que todavía hay un largo camino hacia la alfabetización mediática en términos de la IA generativa, especialmente en lo que respecta a las cuestiones de noticias falsas y la comprensión del riesgo de robo de imagen asociado con los *deepfake*. Aquí, el arte puede desempeñar un papel crucial como una herramienta de crítica a través de la expresividad creativa. Al utilizar el espacio del arte, se abre un margen de libertad donde se puede explorar la apropiación de imágenes, jugar con sus límites y, quizás, descubrir cómo la IA generativa podrá ser un medio para reforzar nuestra libertad democrática, inclusión y expresión.

Tal vez a través del arte podamos jugar con estas (pseudo)realidades generadas por IA, desafiando la percepción y apoyando el desarrollo de una alfabetización mediática más adecuada para el futuro de la educación, en una escuela que, además de enseñar, pueda aprender y desarrollarse junto con su comunidad. Es fundamental que sigamos promoviendo actividades que no solo enseñen a utilizar estas tecnologías, sino que también fomenten una comprensión crítica de sus implicaciones, preparando a los estudiantes (y a nosotros mismos como docentes) para un mundo en el que la IA será cada vez más predominante y poderosa.

Además, la experiencia del proyecto subraya la necesidad de incorporar discusiones éticas acerca de la IA en los planes de estudio en todas las asignaturas del currículo, algo ya preconizado por el Referencial de la Educación en los Medios (MEC, 2014), asegurando que los estudiantes no solo sean consumidores de tecnología, sino también críticos y creadores activos y responsables. Al final, el verdadero valor de la IA en la educación radica no solo en su capacidad para ayudarnos a generar textos, imágenes y sonidos antes inimaginables, sino también en su potencial para empoderar a los estudiantes a cuestionar, innovar y expresarse creativamente en un mundo cada vez más complejo.

## FUENTES REFERENCIALES

Acaso, M., y Megías, C. (2018). *Art thinking*. Paidós.

Comisión Europea. (2022). Directrices éticas sobre el uso de la inteligencia artificial (IA) y los datos en la educación y formación para los educadores. Dirección General de Educación, Juventud, Deporte y Cultura. Oficina de Publicaciones de la Unión Europea. <https://doi.org/10.2766/898>

Deepfakenow. (2023). Understanding The Digital Mirage: Deepfakes Vs Shallowfakes. Recuperado de <https://deepfakenow.com/what-is-the-difference-between-a-deepfake-and-shallowfake/>

Depoorter, D. (2022). The Follower: Using open cameras and AI to find how an Instagram photo is taken. [media art] YouTube and Instagram. Retrieved from: <https://driesdepoorter.be/thefollower/>



- Glover, E. (2024, 14 de marzo). What Are AI Hallucinations? Generative AI models give misleading information sometimes. Here's why. Recuperado de <https://builtin.com/artificial-intelligence/ai-hallucination>
- Guarnera, L., Giudice, O. y Battiato, S. (2020). Fighting Deepfake by Exposing the Convolutional Traces on Images. *IEEE Access*, 8, 165085-165098. <https://doi.org/10.1109/ACCESS.2020.3023037>
- Harrell, D. F., Glick, J., Sahu, R., Herrera, N., Slavina, K. y Nalezty, C. (2021). *Media literacy in the age of deepfakes* [Course module]. MIT Center for Advanced Virtuality.
- Hopkins, D. (2014). *A Teacher's Guide to Classroom Research*. (5th ed.). Open University Press.
- Lessig, L. (2009). *Remix: making art and commerce in the hybrid economy*. Penguin.
- MEC. (2014). *Media Education Guidance*. Retrieved from [https://www.dge.mec.pt/sites/default/files/ECidadania/Referenciais/media\\_education\\_guidance\\_dge\\_pt.pdf](https://www.dge.mec.pt/sites/default/files/ECidadania/Referenciais/media_education_guidance_dge_pt.pdf)
- Payne, L. (2024). Deepfake – AI-generated synthetic media. En *Encyclopaedia Britannica* (en línea). <https://www.britannica.com/technology/deepfake>
- Seow, J., Lim, M., Phan, R. C. W. y Liu, J. (2022). A comprehensive overview of Deepfake: generation, detection, datasets, and opportunities. *Neurocomputing*, 513, 351-371. <https://doi.org/10.1016/j.neucom.2022.09.135>
- Steyerl, H. (2019). *The City of Broken Windows* [Video Installation]. Castello di Rivoli. Retrieved from: <https://www.castellodirivoli.org/en/mostra/hito-steyerl-the-city-of-broken-windows/>
- Weikmann, T. y Lecheler, S. (2022). Visual disinformation in a digital age: A literature synthesis and research agenda. *New Media & Society*, December 2022. <https://doi.org/10.1177/14614448221141648>
- Yang, L. (2017). *Delusional Mandala* [Video]. Retrieved from: <http://luyang.asia/2017/02/15/cartoon-brew-luyang-delusional-mandala-by-luyang-nsfw/>