

Trabalho filosófico

Filosofia, inteligência artificial e literacia mediática num mundo em crise

Democracia, desinformação e consumo mediático
Isabel Bernardo

Pixabay

Organização

Associação de Professores de Filosofia

Ficha técnica

Autor: Isabel Bernardo

Título: Democracia, desinformação e consume mediático

Data: 10.11.2025

Comunicação realizada no âmbito do ciclo temático “O trabalho filosófico 2025: Filosofia, Inteligência Artificial num mundo em crise”

Organização: Associação de Professores de Filosofia

Edição: Associação de Professores de Filosofia

Comunicação sob licença Creative Commons 4.0 (atribuir, sem derivações ou uso comercial)



Nesta intervenção parte-se de três ideias-chave que se encadeiam: a curadoria algorítmica reorganiza a esfera pública ao decidir o que vemos e em que ordem; a desinformação sintética acelera a erosão da confiança, inundando o espaço comum com conteúdos plausíveis mas falsos; e os nossos padrões de consumo mediático — acelerados, personalizados e fragmentados — moldam crenças e, com elas, a participação cívica.

Tendo em conta este enquadramento, questiona-se de que modo a disciplina de Filosofia do ensino secundário pode mobilizar conceitos e competências que permitam desenvolver nos alunos capacidades em literacia mediática e o que isso implica, tanto do ponto de vista das políticas de escola, do ponto de vista pedagógico e didático.



SUMÁRIO

1 Do consumo à literacia mediática

2 Do algoritmo à modificação dos comportamentos: o *time watch* e o caso dos *heartbeats*

3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

4 As políticas de escola e o papel da disciplina de Filosofia: alguns exemplos a partir das aprendizagens essenciais da disciplina de Filosofia

5 Referências bibliográficas

6 Mini glossário



1 Do consumo à literacia mediática

O consumo mediático é simplesmente **a ação e o volume**, ou seja, é a soma de todas as nossas interações com conteúdos mediáticos.

Inclui...

O quê... Os tipos de conteúdo (vídeos no YouTube, *reels* no Instagram, notícias, filmes na Netflix, *podcasts*, música, *memes*).

Como... as meios de acesso usados (smartphone, televisão, computador, rádio).

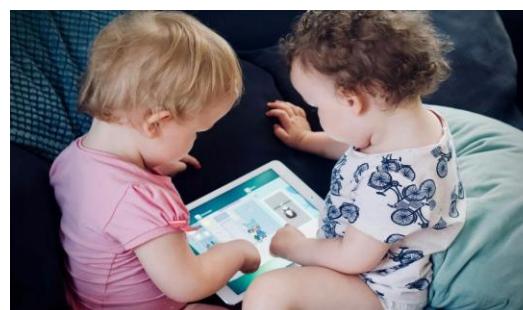
Quanto... a frequência e duração (ex: "3 horas de *screen time* por dia", "ver 50 *reels* seguidos").



Fonte: autor desconhecido, imagem licenciada ao abrigo da [CC BY-SA-NC](https://creativecommons.org/licenses/by-sa/4.0/)



Fonte: imagem gerada por IA



Fonte: autor desconhecido, imagem licenciada ao abrigo da [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/)



Fonte: autor desconhecido, imagem licenciada ao abrigo da [CC BY-ND](https://creativecommons.org/licenses/by-nd/4.0/)



Fonte: autor desconhecido, imagem licenciada ao abrigo da [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/)

1 Do consumo à literacia mediática

Tabela 1 – Principais fontes de consumo mediático para assuntos sociais e políticos na União Europeia para jovens de 16-30

Indicador	Valor
Redes sociais	42%
Televisão	39%
Sites/plataformas de notícias	26%
Amigos/família/colegas	25%
Plataformas de vídeo (YouTube, etc.)	23%
Rádio	16%
Jornais/revistas em papel	12%
Professores (escola/universidade)	11%
Podcasts	11%

Fonte: European Parliament — *Youth Survey 2024 (Flash Eurobarometer)*.

Social media platforms used for information on political and social issues

When respondents were asked which social media and platforms they use to obtain information on political and social issues, **Instagram** emerges as the most commonly used one (47%), followed by **TikTok** (39%) and **YouTube** (37%).

Lower shares obtain information from Facebook (27%) and X (Twitter) (21%). WhatsApp is used for information on political and social issues by 16% of respondents. A small proportion of respondents (8%) indicate that they do not use social media to find information on political and social issues.

European Parliament. (2025). *Youth Survey 2024 (Flash Eurobarometer)*. https://cdn.edupedu.ro/wp-content/uploads/2025/02/Youth_Survey_2024_FL013EP_report_en.pdf



1 Do consumo à literacia mediática

Tabela 2 – Principais fontes de consumo mediático para assuntos sociais e políticos para jovens portugueses de 16-30

Indicador	Valor
Televisão	53%
Redes sociais (ex.: Instagram, TikTok)	40%
Sites/plataformas de notícias	31%
Amigos/família/colegas	25%
Plataformas de vídeo (ex.: YouTube)	20%
Rádio	16%
Jornais/revistas em papel	13%
Professores (escola/universidade)	11%
Podcasts	13%
Publicidade/partidos	11%
Organizações/educadores de juventude	7%

Fonte: Youth Survey 2024 (Flash Eurobarometer – European Parliament), Q7 (Portugal).

Young women are more likely than young men to obtain information on political and social issues from social media platforms (e.g., Instagram, TikTok, etc.) (46% vs 38%), TV (41% vs 38%), and friends, family and colleagues (29% vs 22%). Conversely, young men are more likely to do so from video platforms (e.g., YouTube, etc.) (28% vs 18%).

The youngest group of respondents (aged 16-18) are more likely than their older counterparts to obtain information on political and social issues from social media platforms (45% vs 39% of 25-30 year-olds), friends, family or colleagues (29% vs 23%) and school teachers or university lecturers (18% vs 8%).

Respondents aged 25-30, in turn, are more likely to obtain information from online press and/or news platforms (30% vs 21% of 16-18 year-olds), radio (19% vs 13%), printed newspapers and magazines (13% vs 10%), and podcasts (12% vs 9%).

European Parliament. (2025). Youth Survey 2024 (Flash Eurobarometer). https://cdn.edupedu.ro/wp-content/uploads/2025/02/Youth_Survey_2024_FL013EP_report_en.pdf

1 Do consumo à literacia mediática

A literacia mediática é um conjunto de **capacidades e disposições de pensamento crítico** aplicadas aos media.

É a capacidade de...

Aceder: encontrar e aceder a informação de diferentes fontes.

Analisar: desconstruir uma mensagem mediática, o que envolve fazer perguntas-chave:

- **Quem** criou esta mensagem?
- **Porquê** é que esta mensagem foi criada? (Para informar? Para persuadir? Para vender algo? Para provocar raiva?)
- **Quais** são as técnicas criativas usadas para prender a minha atenção? (Música dramática, edição rápida, linguagem emocional, etc.)
- **O que** é que foi deixado *de fora* desta mensagem?

Avaliar: julgar a credibilidade, fiabilidade e o viés (parcialidade) da informação. É aqui que se distingue facto de opinião, propaganda de notícia e informação verificada de desinformação.

Criar: Produzir os seus próprios conteúdos mediáticos de forma ética e responsável (ex: saber como escrever um comentário construtivo em vez de um ataque, ou como partilhar informação de forma responsável).



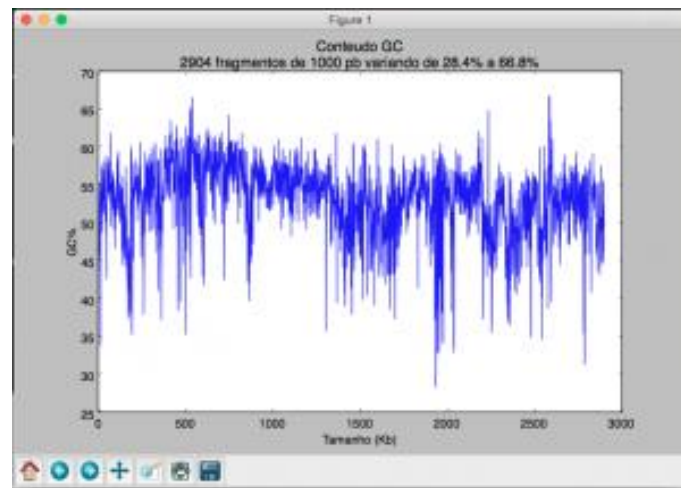
1 Do consumo à literacia mediática

Quadro 1 – Do consumo mediático à literacia mediática: do receber ao avaliar

Característica	Consumo mediático	Literacia mediática
Foco	A ação de ver/ouvir/ler	A capacidade de analisar/avaliar
Pergunta-Chave	"O que estou a ver?"	"Porquê é que estou a ver isto?"
Resultado	Informação, entretenimento, distração	Compreensão, discernimento, ceticismo
Relação com o "Watch Time"	É o ato que gera "watch time"	Questiona ou lidar de modo adequado com o "watch time".



2 Do algoritmo à modificação dos comportamentos: o *time watch* e o caso dos *heartbeats*



Fonte: autor desconhecido. Imagem licenciada ao abrigo da [CC BY-NC](#)

Medição do tempo de visualização (time watch)

Eventos do leitor

Quando vê um vídeo, o leitor de vídeo no seu telemóvel ou navegador está constantemente a registar ações. Cada "play", "pause", "skip" (saltar para a frente), "rewind" (retroceder) e até mesmo a mudança de volume ou qualidade de vídeo são "eventos" registados.

"Heartbeats" (batimentos cardíacos):

Este é o mecanismo principal. Para medir o tempo de visualização contínuo, o leitor de vídeo envia um pequeno sinal (um "ping" ou "heartbeat") aos servidores da plataforma em intervalos regulares (por exemplo, a cada 5, 10 ou 30 segundos).

Registo de fim

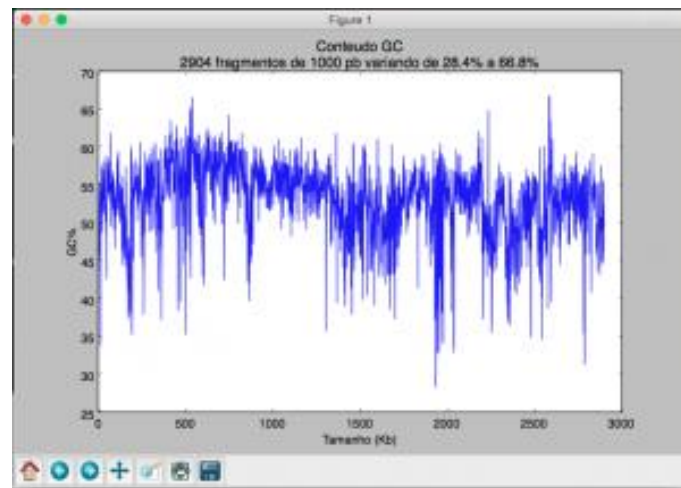
Quando fecha o vídeo, fecha a aplicação, ou o vídeo termina, um evento final é enviado para o servidor, registando o ponto de paragem exato.

Agregação de dados

Os servidores da empresa (ex: Google, Meta, ByteDance) recebem estes triliões de "pings" de todos os utilizadores e agregam-nos. Eles não calculam apenas o "watch time" por vídeo, mas também por utilizador, por criador, por tópico e para a plataforma como um todo.

Os algoritmos que maximizam o tempo de visualização *watch time* são **sistemas de recomendação** avançados, construídos maioritariamente sobre **Deep Learning** (redes neuronais), que são treinados (via **aprendizagem supervisionada** e por **reforço**) usando o "tempo de visualização" como a principal métrica de sucesso.

2 Do algoritmo à modificação dos comportamentos: o *time watch* e o caso dos *heartbeats*



Fonte: autor desconhecido. Imagem licenciada ao abrigo da [CC BY-NC](#)

O “rei” do algoritmo de recomendação

O "watch time" é o principal indicador para o algoritmo de que um conteúdo é **valioso e cativante**: se as pessoas passam mais tempo a ver um determinado vídeo, ele deve ser bom. Se um vídeo é bom, devemos mostrá-lo a mais pessoas para que elas também passem mais tempo na nossa plataforma.

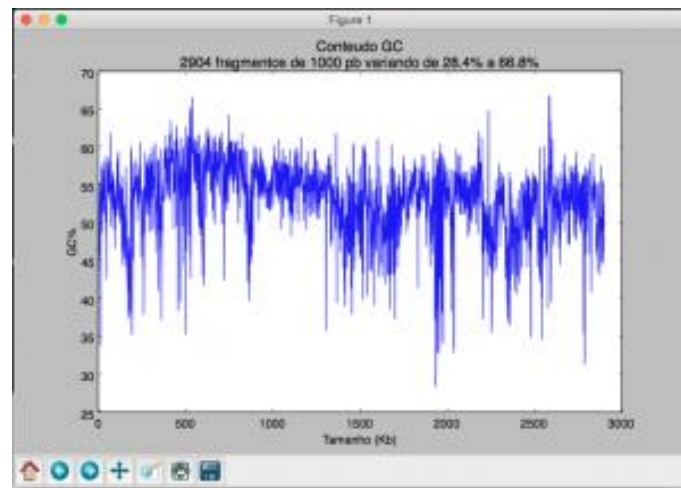
É por isso que um vídeo com 1 milhão de visualizações, mas com um "watch time" médio de apenas 10 segundos, é considerado pelo algoritmo como um "fracasso" (provavelmente *clickbait*). Em contraste, um vídeo com 100.000 visualizações, mas com um "watch time" médio de 8 minutos, é um "sucesso" e será recomendado massivamente.

O motor da monetização e publicidade: mais “watch time” é igual a mais anúncios

O "watch time" está diretamente ligado à receita: quanto mais tempo passar na plataforma, mais oportunidades a plataforma tem de lhe mostrar anúncios.

Os algoritmos que maximizam o tempo de visualização *watch time* são **sistemas de recomendação** avançados, construídos majoritariamente sobre **Deep Learning** (redes neurais), que são treinados (via **aprendizagem supervisionada** e por **reforço**) usando o "tempo de visualização" como a principal métrica de sucesso.

2 Do algoritmo à modificação dos comportamentos: o *time watch* e o caso dos *heartbeats*



Fonte: autor desconhecido. Imagem licenciada ao abrigo da [CC BY-NC](#)

As Métricas-Chave (mais do que apenas “minutos”)

Tempo de visualização absoluto: o número total de horas que as pessoas passaram a ver os seus vídeos (é nisto que o YouTube se foca, favorecendo conteúdo longo).

Retenção de público: a percentagem média do seu vídeo que as pessoas veem (um vídeo de 10 minutos com 50% de retenção (5 min) é excelente).

Taxa de conclusão: como os vídeos são curtos, o algoritmo quer saber quantos utilizadores viram o vídeo *inteiro* (100%) - crucial para o **TikTok e Reels**.

Tempo de sessão: esta é a métrica *secreta* e mais importante. O algoritmo não quer saber apenas se as pessoas gostaram do *seu* vídeo; quer saber se o seu vídeo fez com que as pessoas **ficassem na plataforma a ver mais vídeos**. Se o seu vídeo faz o utilizador fechar a aplicação, isso é um sinal negativo, mesmo que o tenham visto até ao fim.

Os algoritmos que maximizam o tempo de visualização *watch time* são **sistemas de recomendação** avançados, construídos maioritariamente sobre **Deep Learning** (redes neuronais), que são treinados (via **aprendizagem supervisionada** e por **reforço**) usando o "tempo de visualização" como a principal métrica de sucesso.

3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

O papel dos média em democracia

Informar com veracidade e contexto: permite decisões eleitorais e cívicas bem-fundamentadas.

Esfera pública e deliberação: dão palco a debates, ideias concorrentes e contra-narrativas.

Accountability: investigam abusos, corrupção e ineficiência.

Definição da agenda e enquadramento: selecionam temas e moldam o modo como o público os entende.

Pluralismo e inclusão: visibilizam minorias e ampliam vozes fora do centro.

Educação cívica: explicam políticas públicas, processos eleitorais e direitos.

O que reforça a contribuição democrática

Independência editorial e modelos de financiamento sustentáveis (incl. serviço público robusto).

Transparência: propriedade dos média, regras de publicidade política, explicabilidade algorítmica.

Literacia mediática e de IA: verificar fontes, rastrear provas, identificar vieses.

Regulação inteligente: proteção da liberdade de expressão + obrigações de transparência/contas.

Diversidade e proximidade: ecossistema com média nacionais, regionais e comunitários.

Riscos para a democracia (se mal regulados ou capturados)

Concentração e captura: menos diversidade, maior influência de interesses particulares.

Desinformação e propaganda (incluindo IA generativa): erosão da confiança pública.

Curadoria algorítmica/bolhas: polarização, microdirecionamento opaco e manipulação da atenção.

Infotainment e corrida ao clique: superficialidade, mais emoções e menos evidências.



3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

Amplificação de conteúdo divisivo e emocional

O conteúdo que provoca emoções fortes (raiva, medo, indignação, validação) gera, de longe, o maior "watch time" e engajamento (comentários, partilhas).

Impacto eleitoral

Uma análise política de 10 minutos sobre política fiscal terá, quase sempre, um desempenho algorítmico pior do que um clip de 2 minutos de um candidato a "humilhar" um oponente ou um vídeo de teoria da conspiração. O algoritmo promove o que choca e divide, não o que informa de forma equilibrada, distorcendo a perceção pública sobre o que são os temas mais importantes.

A criação de "câmaras de eco"

O "watch time" é um sinal de preferência. Se assiste a um vídeo que critica o "Candidato A", o algoritmo aprende que gosta desse tipo de conteúdo. Para maximizar o seu "tempo de sessão", ele irá sugerir-lhe mais e mais vídeos que criticam o "Candidato A".

Impacto eleitoral

O eleitor deixa de ser exposto a pontos de vista divergentes. A sua própria visão política é constantemente reforçada, enquanto a oposição é apresentada como irracional ou mal-intencionada. Isto leva a uma **polarização extrema**, onde eleitores de lados opostos não conseguem sequer concordar sobre factos básicos.

A vantagem estrutural da desinformação

A desinformação (fake news) é, muitas vezes, *criada* especificamente para ser mais chocante e emocionalmente cativante do que a realidade.

Impacto eleitoral

Uma notícia falsa, mas bombástica (ex: "Candidato A apanhado em grande escândalo"), gera um "watch time" e partilhas imensas. O algoritmo promove-a massivamente. Mesmo que seja desmentida horas ou dias depois, o vídeo de verificação de factos (*fact-checking*) é geralmente mais "aborrecido" e tem um "watch time" inferior. A desinformação espalha-se mais rápido e mais longe do que a verdade, influenciando a opinião do eleitor com base em mentiras.



3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

O Efeito "Toca do Coelho" (Rabbit Hole)

Para manter o "tempo de sessão" (fazer com que continue a ver vídeos), o algoritmo muitas vezes sugere conteúdo que é progressivamente mais extremo.

Impacto eleitoral

Um utilizador pode começar por ver um vídeo sobre um tema legítimo (ex: "preocupações económicas") e, através de recomendações otimizadas pelo "watch time", ser lentamente guiado para conteúdo de franja, teorias da conspiração ou grupos de ódio. Isto pode radicalizar eleitores moderados.

A "gamificação" da Política

Os políticos e as suas equipas de campanha percebem isto. Eles sabem que não ganham a "batalha do algoritmo" com propostas políticas detalhadas, mas sim com "soundbites" (frases de efeito) e momentos criados para viralizar.

Impacto eleitoral

As campanhas focam-se menos em debate de substância e mais em criar conteúdo que "jogue o jogo" do "watch time". A performance mediática e a capacidade de gerar indignação online tornam-se mais importantes do que a competência ou a política pública.

Em suma

O "watch time" é o motor de um sistema que **premeia o extremismo, a desinformação e a divisão**, ao mesmo tempo que marginaliza a moderação e o debate complexo.

O eleitor fica preso numa bolha de informação personalizada, não para o informar, mas para o manter "agarrado" ao ecrã o máximo de tempo possível.

Brady, W. J., Wills, J. A., Jost, J. T., & Crockett, M. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313–7318. <https://doi.org/10.1073/pnas.1618723114>

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira Jr., W. (2020). Auditing radicalization pathways on YouTube. In *FACCT '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 413–424). Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372879>

Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>

3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

Quadro 2 – Dados no âmbito das eleições em Portugal (2020 e 2025) e relativos à marcação da agenda pública a partir das redes sociais

Métrica	Valor	Período / Âmbito	Fonte
Denúncias via WhatsApp recebidas pela CNE/MediaLab Iscte (Europeias 2024)	19 denúncias, 13 analisadas	Pré-campanha + campanha (Portugal)	CNE / MediaLab Iscte – Relatório-síntese (PDF)
Temas de desinformação mais recorrentes (Portugal)	Imigração em alta; casos de corrupção; 1.º deepfake de líder político PT	Europeias 2024	CNE / MediaLab Iscte – Relatório-síntese (PDF)
Interações do caso 'Tiros em Famalicão' (exemplo de amplificação)	Post CHEGA (FB): 3.013/4.796 (60,6%); AV (X): 5.028/8.548 (58,8%)	Legislativas 2024 (exemplo comparativo)	CNE / MediaLab Iscte – Relatório-síntese (PDF)
Crescimento de publicações sobre imigração no Facebook (Portugal)	>200% (2024 → 2025)	Comparação entre Legislativas 2024 e 2025	OberCom / IBERIFIER – Legislativas 2025 (Zenodo)
Plataforma com mais interações/visualizações para partidos/candidatos	Facebook (X como rede de discussão política; crescimento de TikTok)	Legislativas 2025	OberCom – Página do relatório

Nota complementar

O caso Famalicão - durante a pré-campanha de 2024 circulou nas redes a alegação de que teriam ocorrido “tiros” contra a comitiva do CHEGA em Famalicão. Horas depois, a informação foi corrigida/desmentida (não houve tiros). **Métricas de interação** (exemplos reportados): Post do CHEGA (Facebook): 3.013/4.796 (60,6%). Post de André Ventura (X): 5.028/8.548 (58,8%). **Efeito megafone** - atores políticos com grandes audiências funcionam como multiplicadores.

3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

Quadro 3 – Estratégias das plataformas para controlar a desinformação

Estratégia	Objetivo	O que fazem	Como implementam	Exemplos típicos	Impacto no “watch time”/alcance	Desafios / Problemas contínuos
Remover (moderação de conteúdo)	Eliminar conteúdos que violam regras explícitas	Aplicam “Políticas da Comunidade”/“Padrões” e removem conteúdos ilegais/perigosos	IA para deteção em massa + revisão por moderadores humanos; reforço em períodos eleitorais	Desinformação sobre como/quando/onde votar; incitação à violência; deepfakes enganadores; assédio direcionado	Reduz imediatamente o alcance (conteúdo desaparece), mas pode gerar reação pública	Escala (impossível rever tudo), rapidez (desinformação viraliza antes da remoção), zona cinzenta (sátira vs. dano), acusações de censura (erosão de confiança)
Reduzir (ajustes algorítmicos)	Diminuir a difusão de conteúdo “limítrofe” sem o remover	Identificam conteúdo de baixa qualidade/sensacionalista que não viola regras	“Downranking”: não recomendar; manter acessível via pesquisa; otimização por “qualidade/satisfação”	Menor promoção de teorias da conspiração; priorização de fontes “de autoridade”	Diminui o tempo de visionamento de conteúdos borderline; desloca métricas para “satisfação” e “qualidade”	Transparência limitada sobre critérios, risco de viés algorítmico, conflito de interesses (reduções podem afetar receita), adaptação dos maus atores a sinais do algoritmo
Informar (contexto e verificação)	Inocular utilizadores com contexto fiável	Adicionam etiquetas, painéis informativos, avisos de verificação	Parcerias com fact-checkers; redução de distribuição de conteúdos “Falso/Parcialmente falso”; painéis sobre eleições/saúde; etiquetas de “média estatal”	Avisos com ligações para comissões eleitorais, ONU, artigos de verificação independentes	Pode reduzir cliques no sensacionalismo e aumentar consumo informado, com impacto variável no “watch time”	Reação por vezes lenta, sobrecarga de rótulos (fadiga do utilizador), disputa sobre autoridade das fontes, criadores contornam com linguagem codificada (“dog whistles”, “só a fazer perguntas”)

Problemas transversais: reação tardia a campanhas já virais; tensão entre segurança e um modelo de negócio centrado no engajamento; e rápida adaptação dos disseminadores de desinformação, que evitam palavras-chave e exploram zonas cinzentas.

3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

Quadro 4 – Esforço reportado das plataformas no controlo da desinformação em eleições na EU entre 2023 e 2024 (com exceção do Telegram)

Plataforma	Indicador	Valor	Janela temporal	Como interpretar	Fonte (título)	Fonte (link)
X (Twitter)	Contas associadas a redes de desinformação (amostra UE)	≈ 50.000 contas	2024	Contagem de contas detectadas em estudos; não é o total de bots da plataforma.	The Guardian – 12 jul 2024	https://www.theguardian.com/world/article/2024/jul/12/disinformation-networks-social-media-x-france-germany-italy-eu-elections
X (Twitter)	Perfis com foto de perfil gerada por IA (proxy de perfis sintéticos)	≈ 0,052% de ~15 milhões de perfis analisados	2024–2025	Percentagem mínima observada; sugere uso mensurável de perfis sintéticos, não o nº total de bots.	Estudos académicos/monitorização (sintetizado)	https://www.theguardian.com/world/article/2024/jul/12/disinformation-networks-social-media-x-france-germany-italy-eu-elections
Meta (Facebook Instagram)	Redes de comportamento inautêntico coordenado removidas	Centenas a milhares de contas por operação (ex.: 1.020 FB + 711 IG em Q4 2023)	2023–2024	Mede remoções por operações específicas; indica escala de redes coordenadas, não censo total.	Relatórios de ameaças/aplicação de políticas (Meta)	https://transparency.meta.com
Meta (Facebook)	% estimada de contas falsas nos utilizadores mensais	≈ 3–4% (estimativa da própria Meta)	Relatórios de transparência 2024	Estimativa global de contas falsas; nem todas geram conteúdos sintéticos.	Meta (Transparency Center)	https://transparency.meta.com
TikTok	Vídeos com media sintética removidos	≈ 51.618 vídeos removidos (2024)	2024 (H2) – reportado em 2025	Remoções de conteúdo sintético; um bom proxy para a escala do fenómeno.	TikTok – Relatórios de Transparência (EU Code of Practice)	https://www.tiktok.com/transparency/pt-pt
TikTok	Contas e 'likes' falsos removidos (automatização/engajamento falso)	≈ 10 milhões de contas falsas e 460 milhões de 'likes' falsos (2024)	2024 (H2) – reportado em 2025	Indica esforço anti-manipulação; não é contagem de bots políticos, mas mostra escala.	TikTok – Relatórios de Transparência	https://www.tiktok.com/transparency/pt-pt
Telegram (bots)	Bots que geram/espalham deepfakes (escala de utilização)	≥ 50 bots com > 4 milhões de utilizadores/mês	2024	Mostra escala de bots de <i>deepfake</i> ; não é focado em política, mas em capacidade técnica.	Levantamentos jornalísticos (investigação)	https://www.theguardian.com/technology

H2 (half two) – segunda metade do ano.

3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

Quadro 5 – Variação na aplicação das regras de verificação e controlo da desinformação à escala global e origem geopolítica das empresas tecnológicas

Pergunta	Resposta	Principais fatores / obrigações	Exemplos ilustrativos	Implicações
As medidas são aplicadas igualmente em todo o planeta?	Não. A aplicação varia muito por país/mercado.	<p>Leis locais (geo-blocking): plataformas cumprem a lei de cada país para manter acesso ao mercado.</p> <p>Recursos de moderação: investimento concentra-se nas línguas/mercados mais lucrativos (IA + moderadores humanos).</p> <p>Contexto cultural: interpretações locais de discurso de ódio, nudez, sátira, etc.</p>	<p>Alemanha/França: símbolos nazis/negação do Holocausto são ilegais → conteúdo bloqueado localmente; pode permanecer visível nos EUA.</p> <p>Turquia/Índia: pedidos frequentes de remoção rápida por “segurança nacional” ou crítica a governantes; plataformas muitas vezes cumprem.</p> <p>Mianmar/Etiópia: equipas menores → discurso de ódio/desinformação permanecem mais tempo, com riscos fora do ecrã.</p>	Internet não é uniforme. O mesmo conteúdo pode ser visível num país e bloqueado noutro. Mercados com menos recursos têm maior latência na moderação e maior risco de danos.
EUA vs. China: têm o mesmo “código de conduta”?	Não. Filosofias e obrigações são opostas.	<p>Plataformas dos EUA (Meta, Google/YouTube, X): empresas privadas, moldadas pela cultura legal dos EUA (liberdade de expressão). Equilibram: segurança do utilizador, exigências estatais noutros países, pressão de anunciantes.</p> <p>Plataformas da China: no mercado interno operam sob controlo/censura do Estado (PCC). Fora da China (p.ex., TikTok), afirmam operações separadas e centros de dados locais, mas enfrentam suspeitas devido a leis chinesas de segurança (obrigação de cooperação com inteligência).</p>	<p>Douyin (China): remoção sistemática de críticas ao governo/temas sensíveis (p.ex., Tiananmen).</p> <p>TikTok (global): controvérsia sobre dados e possível influência algorítmica; projetos como “Project Clover” na Europa para separar operações.</p>	EUA (teórico): primazia da livre expressão vs. moderação responsável. China (comprovado internamente): estabilidade/controlo estatal. No exterior, persistem receios de acesso a dados e influência no <i>feed</i> .
As redes são obrigadas a moderar consoante as regras locais?	Sim. Princípio da soberania nacional.	<p>Para operar num país, têm de cumprir as leis locais; sanções variam de multas a bloqueio total.</p> <p>UE – Digital Services Act (DSA): remove conteúdo ilegal rapidamente; transparência algorítmica; direito de recurso; coimas até 6% da receita global por incumprimento.</p>	<p>Bloqueios: Twitter/X na China; Facebook na Rússia (exemplos de restrição/banimento).</p> <p>DSA: obriga a sistemas de mitigação de riscos, auditorias e acesso a dados para investigação (resumo).</p>	<p>Mosaico regulatório: <i>watch time</i> é motor económico, mas as “regras de trânsito” variam por jurisdição.</p> <p>Plataformas adaptam políticas, fluxos de moderação e engenharia de produto a cada mercado.</p>

3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

Quadro 6 – Etapas de *fact-checking* manual

Etapas	O que acontece?	Quem faz?	Como é feito / Ferramentas?	Qual o resultado direto?
1. Detecção	Conteúdos potencialmente falsos são sinalizados.	IA da plataforma; utilizadores; jornalistas/parceiros.	Algoritmos (picos anómalos, links de sites de <i>fake news</i>); denúncias; monitorização ativa de narrativas virais.	Item entra no fluxo de trabalho.
2. “Dashboard”	Conteúdos sinalizados entram numa fila partilhada.	Jornalistas parceiros (p.ex., Polígrafo, LUSA).	Acesso a um painel da Meta com lista priorizada de <i>posts</i> /vídeos/imagens.	Seleção do item a verificar.
3. Verificação	Trabalho jornalístico tradicional.	Equipas de <i>fact-checking</i> .	Contacto com fontes; consulta a peritos; verificação de dados oficiais (PORDATA, INE...); pesquisa inversa de imagem/contexto original.	Artigo/relatório de verificação (<i>fact-check</i>).
4. Classificação	Aplicação de etiqueta ao post original.	Jornalista.	Escolha da etiqueta conforme a evidência.	Etiquetas: Falso; Parcialmente Falso; Contexto em Falta; Sátira; Verdadeiro.
5. Ação da Plataforma	Penalização e avisos são ativados (quando falso/parcialment e falso).	Plataforma.	Algoritmo reduz alcance; overlay de aviso; notificações a quem partilhou; penalização de páginas reincidentes.	Alcance pode cair ≥80%; exige clique para ver; alertas com link para o <i>fact-check</i> .

Quadro 7 – Desafios de *fact-checking* manual

Tema	O que significa na prática	Consequência
Não é remoção (regra geral)	Prefere-se “reduzir e informar” a apagar (salvo outras violações, p.ex., ódio).	Menor acusação de censura, mas o conteúdo permanece acessível.
Escala	Poucos jornalistas vs. volume massivo de <i>posts</i> .	Só uma fração mínima é verificada.
Velocidade	Desinformação viraliza em horas; verificação rigorosa pode levar um dia.	Muitas correções chegam tarde, com dano já feito.
Conteúdo efémero/fechado	<i>Stories</i> , grupos privados e DMs (p.ex., WhatsApp) são difíceis de auditar.	Baixa cobertura do que circula fora do espaço público.

DM = Direct Messages / Mensagens diretas

3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

Quadro 8 – Impactos da IA generativa na criação de desinformação e impacto no *fact-checking*

Dimensão	Antes da IA	Com IA generativa	Impacto no <i>fact-checking</i>	Exemplos
Escala e velocidade (“tsunami” de conteúdo)	Produção de conteúdos falsos exigia tempo, dinheiro e perícia.	Qualquer pessoa gera texto, imagem ou áudio em segundos e quase sem custo.	Volume “industrial” torna impossível acompanhar manualmente.	Imagens/artigos/áudio gerados por modelos (p.ex., Midjourney, LLMs).
Hiper-realismo (erosão da prova)	Análise forense básica (sombas, pixels) muitas vezes bastava.	<i>Deepfakes</i> (sobretudo voz/vídeo) quase indistinguíveis a olho nu e por <i>software</i> .	A prova audiovisual perde força probatória; aumenta a incerteza.	Clonagem de voz a partir de segundos de áudio; vídeos “plausíveis”.
“Dividendo do mentiroso”	Conteúdos autênticos eram, em regra, reconhecidos como tais.	“Tudo pode ser <i>deepfake</i> ”.	Agentes podem negar provas reais dizendo “é IA”; dificulta a responsabilização.	Políticos a rejeitar vídeos/áudios genuínos como “falsos”.
Corrida da deteção	Ferramentas limitadas mas relativamente estáveis.	Detetores ficam desatualizados quando surge novo modelo; falsos positivos/negativos.	Estratégia centrada só na deteção é frágil; exige abordagens de proveniência e literacia.	Detetores de texto/áudio com acurácia inconsistente.
Astroturfing e comentários falsos	Manipulação exigia equipas e contas coordenadas.	Geração massiva de comentários “únicos” com IA.	Distorce perceções de apoio público; difícil distinguir de sentimento genuíno.	Campanhas coordenadas com posts/comentários sintéticos.

Astroturfing — é a criação artificial de uma aparência de apoio popular (“grassroots”) a uma causa, produto ou posição política. Organizações, campanhas ou empresas disfarçam-se de cidadãos comuns ou grupos locais para influenciar opinião pública, media e decisores, deturpando o consenso público, pressionando jornalistas e regulações, e pode contaminar consultas públicas ou eleições.



3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

Quadro 9 – Tentativas de manter a integridade epistémica de imagem e voz pelas companhias tecnológicas e desafios dos modelos *open-source*

Tópico	O que é	Como funciona	Exemplos	Forças	Limitações / Riscos
Marcas de água digitais (watermarking)	Sinal invisível embutido no conteúdo gerado por IA	Inserção de padrões/assinaturas no momento da geração para posterior deteção por “scanner”	SynthID (Google DeepMind), incluindo SynthID Detector para imagens	Transparência; verificação automatizável em larga escala	Fragilidade a recompressão, <i>screenshots</i> , filtros; adoção voluntária (modelos <i>open-source</i> podem ignorar)
Autenticação / Credenciais de conteúdo (proveniência)	“Certidão de nascimento” digital: registo de captura e de edições	Padrão: metadados assinados (autor, dispositivo, hora, GPS; cadeia de edições) consultáveis	Content Credentials (Adobe, Microsoft, BBC, Nikon, Canon, etc.)	Foca-se no conteúdo real (prova de autenticidade); mais robusto a transformações	Requer <i>hardware/software</i> compatível; adoção setorial ainda em expansão
Modelos ocidentais (OpenAI, Google, Meta)	LLMs (texto) e difusão/imagem (visual) com guardrails	Políticas de uso; bloqueio de <i>prompts</i> (políticos/figuras públicas); lançamento restrito de tecnologias de voz	Recusas em gerar imagens de políticos; Voice Engine com lançamento limitado	Reduz riscos de abuso; alinhamento com marca e requisitos legais	<i>Guardrails</i> podem ser contornados via modelos alternativos; risco de falsos positivos/negativos em detetores
Modelos chineses (Baidu, Alibaba, Tencent)	Limitação orientada por regulação estatal e censura política	“Medidas interinas” exigem adesão a valores socialistas; censura de tópicos sensíveis; etiquetagem obrigatória	Ernie Bot (Baidu) com bloqueios temáticos; requisitos de rotulagem	Cobertura legal clara e exigível; uniformidade de práticas	Censura política; impacto na liberdade de expressão/investigação; portabilidade limitada
Modelos open-source (Stable Diffusion, Llama, etc.)	Pesos locais passíveis de modificação	Execução em <i>hardware</i> próprio; remoção de filtros/guardrails	<i>Forks</i> e UIs comunitárias sem salvaguardas	Inovação rápida; acessibilidade	Vetor principal de <i>deepfakes</i> maliciosos; fora do alcance de políticas corporativas

Forks: cópias (“ramificações”) de projetos open-source — por exemplo, modelos ou pipelines — que a comunidade altera e recompila. Nessas versões, os autores podem remover filtros de segurança, mudar limites de resolução/rosto/voz, ou ativar funções bloqueadas nos lançamentos oficiais.

UIs comunitárias: interfaces gráficas feitas pela comunidade (p.ex., webUIs para geração de imagem/voz/vídeo) que tornam tudo “clique-a-clique”. Muitas não impõem *guardrails* (bloqueio de prompts, listas de termos, telemetria, avisos), ou permitem contornar as limitações com um *toggle* (um interruptor que desliga) ou *plugin* (aplicação que permite contornar os limites de segurança).

3 Média, democracia e o impacto do microdirecionamento assente no *time watch*

Quadro 10 – Regulação e políticas de IA: UE, EUA e China (*deepfakes*, desinformação e limites)

Jurisdição	Filosofia central	Mecanismos principais	Regras para deepfakes/desinformação	Situação atual / Observações
União Europeia – o “Regulador”	Direitos fundamentais e gestão de riscos; confiança como base para adoção (“Efeito de Bruxelas”).	Lei da IA (Regulamento (UE) 2024/1689) – regulação horizontal e preventiva por risco.	Regras de transparência: etiquetagem clara de conteúdos sintéticos e transparência de bots (art. 52).	Implementação faseada; convergência com políticas de proveniência (e.g., credenciais de conteúdo) e DSA/DMA.
Estados Unidos – o “Inovador”	Abordagem pró-inovação e setorial; foco em competitividade e liderança tecnológica.	Compromissos voluntários (2023). Ordem Executiva de 2023 (Biden) orientou agências; regulação setorial por agências norte-americanas	Mais reativas e limitadas pela Primeira Emenda (liberdade de expressão). A FCC (2024) classificou vozes geradas por IA em <i>robocalls</i> como ilegais sem consentimento.	2025 (Trump): EO 14179 “Removing Barriers to American Leadership in AI” revoga/manda rever elementos da EO 14110 (Biden), recentrando em desregulação e liderança; processo em evolução com planos subsequentes.
China – o “Controlador”	Estabilidade social e controlo estatal; alinhamento com diretivas do PCC.	Regras de algoritmos (2022) e de “ <i>deep synthesis</i> ” (2023); licenciamento de modelos; etiquetagem obrigatória de conteúdo gerado por IA.	Censura de tópicos sensíveis; proibição rígida de <i>deepfakes</i> sobre líderes chineses; forte rotulagem e responsabilização de provedores.	Aplicação célere e prescritiva; compatibilidade internacional limitada; cumprimento exigível pelo Estado.

Robocalls — são chamadas telefónicas **automáticas** feitas por sistemas que discam números e reproduzem **mensagens gravadas** (ou voz sintetizada), em massa. Servem para marketing, avisos (ex.: companhias aéreas, escolas), cobranças ou propaganda política.

Nota complementar

Em 23/01/2025, a EO 14179 (“Removing Barriers to American Leadership in Artificial Intelligence”) redefiniu a orientação federal, revogando e/ou mandando rever ações decorrentes da EO 14110 (30/10/2023, Biden). A nova linha enfatiza desregulação e liderança; medidas setoriais como a decisão da FCC (08/02/2024) sobre robocalls com voz gerada por IA mantêm-se em vigor.

4 As políticas de escola e o papel da disciplina de Filosofia: alguns exemplos a partir das aprendizagens essenciais da disciplina de Filosofia

Epistemologia - Teoria do conhecimento como crença verdadeira e justificada

O ecossistema mediático moderno (otimizado para o *watch time*) confunde deliberadamente "crença" com "conhecimento".

Pergunta: "Como sei que isto é verdade?"

O professor apresenta duas peças de conteúdo sobre o mesmo evento:

Peça A: um *Reel* ou TikTok viral, cheio de música dramática, opiniões fortes e cortes rápidos (otimizado para *watch time*).

Peça B: um artigo da Agência LUSA ou de um verificador de factos (como o Polígrafo), com fontes, dados e uma linguagem neutra.

Debate:

"Ambos geram em nós uma *crença*. Mas qual deles é *conhecimento*?"

"Aplicamos a definição: O TikTok gera uma 'crença'? Sim. É 'verdadeira'? Talvez. É 'justificada'? Onde está a justificação?"

"O algoritmo premeia a *crença* (a reação emocional) ou a *justificação* (a verificação lenta)?"

Epistemologia – Hume e os *deepfakes*

Hume defende que o conhecimento sobre o mundo tem origem empírica.

O professor introduz David Hume. "Todo o nosso conhecimento empírico baseia-se na confiança nos nossos sentidos (visão, audição)." Em seguida, mostra um *deepfake*.

Pergunta: "O que acontece ao conhecimento se os nossos sentidos podem ser sistematicamente enganados pela IA? Em que podemos confiar?"

4 As políticas de escola e o papel da disciplina de Filosofia: alguns exemplos a partir das aprendizagens essenciais da disciplina de Filosofia

A ferramenta central: Lógica e Argumentação

AE relevante e ângulo relevante: Identificação de falácias informais (ex: *ad hominem*, boneco de palha, apelo à emoção, falso dilema, derrapagem).

"Este argumento é forte? É cogente?"

Ação em aula:

"Caça às falácias" em caixas de comentários: O professor projeta uma caixa de comentários de uma notícia política polémica no Facebook ou YouTube.

Trabalho de Grupo: Os alunos, armados com a sua lista de falácias das AE, têm de identificar:

Ad Hominem: (Ataques ao caráter do político, em vez de ao seu argumento).

Apelo à emoção (Pathos): (Quase 100% dos comentários. "Isto é uma vergonha!").

Boneco de palha: (Distorcer o argumento do lado oposto para o ridicularizar).

Discussão de fundo: O professor pergunta: "Porque é que o 'apelo à emoção' é a falácia mais recompensada pelo algoritmo de *watch time*? Um argumento forte ou sólido, teria o mesmo alcance?"

A avaliação da fonte e os argumentos por autoridade

AE relevante e ângulo filosófico: Argumento por autoridade e respetivas falácias. Qual é a autoridade por detrás desta mensagem?

Ação em aula - O professor compara duas fontes sobre saúde (ex: vacinas): um *post* de um *influencer de lifestyle* vs. um comunicado da Direção-Geral da Saúde (DGS).

Debate

"O *influencer* tem um (apelo emocional, confiança) muito elevado. A DGS tem dados e argumentação forte forte. Mas como avaliamos cada um?"

"O algoritmo consegue distinguir a autoridade de quem argumenta? Consegue distinguir o argumento forte de um argumento falacioso?"

O bem comum: O professor pergunta: "Se uma plataforma sabe que a desinformação A é prejudicial ao 'bem comum', mas gera mais *watch time* do que a verdade B, qual deve ela promover?"

4 As políticas de escola e o papel da disciplina de Filosofia: alguns exemplos a partir das aprendizagens essenciais da disciplina de Filosofia

Ética e plataformas digitais

AE relevante e ângulo filosófico: Éticas de Kant e Mill. "Esta ação é eticamente correta? Quais são as suas consequências?"

Ação em Aula:

O dilema do algoritmo

O professor apresenta um dilema moral:

"Você é o engenheiro-chefe do TikTok. Tem dois algoritmos."

Algoritmo Kantiano (deontológico): baseia-se na máxima "Não usarás o utilizador como um meio para um fim"; mostra apenas conteúdo verificado e educativo, respeitando a autonomia do utilizador. *Resultado: O 'watch time' cai 50%. A empresa corre o risco de falir.*

Algoritmo Utilitarista (de Mill): baseia-se em "maximizar a felicidade (neste caso, o *engagement*) para o maior número"; promove vídeos chocantes, polarizadores e viciantes. *Resultado: O 'watch time' dispara. A empresa prospera, mas a polarização social aumenta.*

Debate: "Qual algoritmo devemos escolher? Devemos seguir o dever (Kant) ou as consequências (Mill)? O *watch time* é uma boa medida de 'felicidade' ou 'bem-estar'?"

Referências bibliográficas

- Alali, A., & Theodorakopoulos, G. (2025). Partial fake speech attacks in the real world using deepfake audio. *Journal of Cybersecurity and Privacy*, 5(1), 6. <https://doi.org/10.3390/jcp5010006>
- Avaaz. (2019, 22 de maio). *Far-right networks of deception: Takedown of online disinformation ahead of the EU elections*. <https://avaazimages.avaaz.org/Avaaz%20Report%20Network%20Deception%2020190522.pdf>
- Brady, W. J., Wills, J. A., Jost, J. T., & Crockett, M. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313–7318. <https://pubmed.ncbi.nlm.nih.gov/28652356/>
- Cardoso, G., Moreno, J., Narciso, I., Couraceiro, P., & dos Santos, J. G. B. (2025, 3 de julho). *Legislativas 2025 – Informação e desinformação nas redes sociais*. OberCom – Observatório da Comunicação. <https://doi.org/10.5281/zenodo.15781757>
- ChinaLawTranslate. (2017/2018). *National Intelligence Law of the People's Republic of China (2017, as amended 2018)* [English translation]. <https://www.chinalawtranslate.com/en/national-intelligence-law-of-the-p-r-c-2017/>
- Coalition for Content Provenance and Authenticity. (n.d.). Advancing digital content transparency. C2PA. <https://c2pa.org/>
- Comissão Nacional de Eleições, & MediaLab Iscte. (2024, 9 de julho). *Europeias 2024: Relatório-síntese — Protocolo CNE–MediaLab Iscte para a monitorização e despistagem de desinformação política*. https://www.cne.pt/sites/default/files/dl/eleicoes/2024_pe/docs_geral/relatorio-sintese_eleicoes-europeias-2024_redes-sociais_monitorizacao-desinformacao.pdf
- Cyberspace Administration of China. (2023). *Provisions on the Administration of Deep Synthesis Internet Information Services (English trans.)*. *China Law Translate*. <https://www.chinalawtranslate.com/en/deep-synthesis/>
- Cyberspace Administration of China. (2023, July 10). *Interim Measures for the Management of Generative Artificial Intelligence Services (English trans.)*. *China Law Translate*. <https://www.chinalawtranslate.com/en/generative-ai-interim/>
- European Commission. (2025, 6 de junho). *Report on the 2024 elections to the European Parliament*. https://commission.europa.eu/publications/report-2024-elections-european-parliament_en (PDF: https://commission.europa.eu/document/download/2a7fddb2-e927-4079-92cc-4bb4279e9a46_en?filename=Report+on+the+2024+elections+to+the+European+Parliament.pdf)
- European Commission. (2025, August 22). *The Digital Services Act (DSA)*. European Commission. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en
- European Commission. (2025, October 24). *Commission preliminarily finds TikTok and Meta in breach of certain obligations under the DSA* [Press release]. European Commission. <https://commission.europa.eu/>
- European Digital Media Observatory (EDMO). (2024, 24 de julho). *Final report — Results and outcomes of a community-wide effort (Task Force on the 2024 European Parliament elections)*. <https://edmo.eu/publications/final-report-results-and-outcomes-of-a-community-wide-effort/>
- European Fact-Checking Standards Network. (n.d.). *European Code of Standards for Independent Fact-Checking Organisations*. <https://efcsn.com/code-of-standards/>
- European Union. (2024). *Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. EUR-Lex. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
- Federal Communications Commission. (2024, February 8). *Declaratory ruling: AI-generated voice calls are illegal under the TCPA, absent consent*. <https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal>
- Federal Ministry of Justice (Germany). (n.d.). *Strafgesetzbuch (StGB) § 130 — Incitement to hatred* [English translation]. https://www.gesetze-im-internet.de/englisch_stgb/
- Goldstein, J. A., & Lohn, A. (2024, January 23). *Deepfakes, elections, and shrinking the liar's dividend*. Brennan Center for Justice. <https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend>
- Google DeepMind. (2025, May 20). *SynthID Detector — a new portal to help identify AI-generated content*. Google. <https://blog.google/technology/ai/google-synthid-ai-content-detector/>
- Legifrance. (1990, July 13). *Loi n° 90-615 tendant à réprimer tout acte raciste, antisémite ou xénophobe (Loi Gayssot)*. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000532990>
- Meta Platforms, Inc. (2025). *Community Standards Enforcement — Transparency Center* (síntese de relatórios trimestrais). <https://transparency.meta.com/reports/community-standards-enforcement/> (transparency.meta.com)

Referências bibliográficas

Meta Platforms, Inc. (2025). *Community Standards Enforcement — Transparency Center* (síntese de relatórios trimestrais). <https://transparency.meta.com/reports/community-standards-enforcement/> (transparency.meta.com)

Meta Transparency Center. (2025, April 7). How fact-checking works. Meta. <https://transparency.meta.com/features/how-fact-checking-works/>

Negreiro, M. (2025, July). Children and deepfakes. European Parliamentary Research Service. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775855/EPRS_BRI\(2025\)775855_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775855/EPRS_BRI(2025)775855_EN.pdf)

Poynter Institute/IFCN. (n.d.). IFCN Code of Principles. <https://ifcncodeofprinciples.poynter.org/>

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira Jr., W. (2020). Auditing radicalization pathways on YouTube. In *FACCT '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 413–424). Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372879>

Swenson, A. (2024, February 8). AI-generated voices in robocalls can deceive voters. The FCC just made them illegal. AP News. <https://apnews.com/article/a8292b1371b3764916461f60660b93e6>

The White House. (2023, October 30). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (EO 14110). <https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>

The White House. (2025, January 23). Removing Barriers to American Leadership in Artificial Intelligence (EO 14179). <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>

TikTok. (2025, 25 de março). *TikTok's fifth transparency report: Our progress under the EU disinformation code*. <https://newsroom.tiktok.com/en-eu/tiktoks-fifth-transparency-report-our-progress-under-the-eu-disinformation-code> (newsroom.tiktok.com)

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(26). <https://doi.org/10.1007/s40979-023-00146-z>

X Corp. (2023, agosto). *Civic integrity policy*. X Help Center. <https://help.x.com/en/rules-and-policies/election-integrity-policy>

YouTube. (2019, 25 de janeiro). *Continuing our work to improve recommendations on YouTube*. YouTube Official Blog. <https://blog.youtube/news-and-events/continuing-our-work-to-improve/> blog.youtube



Mini glossário

Algoritmo. Tipos de algoritmo

Um **algoritmo** é um conjunto de instruções passo a passo, finitas e bem definidas, concebidas para realizar uma tarefa específica ou resolver um problema

Os principais tipos de algoritmos de IA usados hoje dividem-se nestas grandes categorias:

1. Aprendizagem Automática (Machine Learning)

Este é o grande pilar da IA moderna.

•Aprendizagem Supervisionada (Supervised Learning)

- **O que é:** O algoritmo aprende a partir de dados que já estão "rotulados" com a resposta correta. É como aprender com um professor que lhe dá exemplos e as respetivas soluções.
- **Exemplos de algoritmos:** Regressão Linear, Árvores de Decisão, Support Vector Machines (SVM).
- **Onde é usado:** Filtros de spam (o email é "spam" ou "não spam"), reconhecimento facial (esta foto é da "Pessoa A"), previsão do tempo.

•Aprendizagem Não Supervisionada (Unsupervised Learning)

- **O que é:** O algoritmo recebe dados "não rotulados" e tem de encontrar padrões ou estruturas por conta própria. É como ser largado numa biblioteca e ter de organizar os livros por tema sem saber os temas de antemão.
- **Exemplos de algoritmos:** K-Means (Clustering), Análise de Componentes Principais (PCA).
- **Onde é usado:** Segmentação de clientes (agrupar clientes com comportamentos de compra semelhantes), deteção de anomalias (detetar fraude bancária).

•Aprendizagem por Reforço (Reinforcement Learning)

- **O que é:** O algoritmo aprende através de "tentativa e erro". Recebe "recompensas" por ações corretas e "punições" por ações erradas, aprendendo a estratégia que maximiza a recompensa.
- **Onde é usado:** Carros autónomos (recompensa por ficar na faixa), IA para jogos (como o AlphaGo, que aprendeu a jogar Go), otimização de sistemas de logística.

2. Aprendizagem Profunda (Deep Learning)

Esta é a área mais conhecida e poderosa da IA atualmente. É uma subárea muito avançada do Machine Learning, que usa algoritmos específicos chamados Redes Neurais Artificiais.

Estas redes são inspiradas no cérebro humano e são "profundas" (deep) porque têm muitas camadas de neurónios artificiais.

•Redes Neurais Convolucionais (CNNs)

- **O que fazem:** São especializadas em processar dados visuais.
- **Onde são usadas:** Reconhecimento de imagens (o que está numa foto), carros autónomos (identificar sinais de trânsito, peões), análise de exames médicos (raios-X, TACs).

•Redes Neurais Recorrentes (RNNs) e Transformers

- **O que fazem:** São especializadas em processar sequências de dados, como texto ou áudio. Os **Transformers** são a arquitetura mais moderna e dominante.
- **Onde são usadas:** É o que permite que assistentes como eu (ChatGPT, Gemini) entendam e gerem linguagem. Também usadas em tradução automática e legendagem de vídeos.

3. Outros Algoritmos de IA (Mais Clássicos)

Embora o Deep Learning domine as notícias, estes algoritmos continuam a ser fundamentais:

•Algoritmos de Busca e Otimização (ex: A*, Minimax)

- Usados para encontrar a melhor solução entre muitas possíveis. O GPS do seu telemóvel usa um algoritmo de busca (como o A*) para encontrar o caminho mais rápido.

•Algoritmos Evolutivos (ex: Algoritmos Genéticos)

- Inspirados na seleção natural, são usados para encontrar soluções ótimas para problemas complexos, "evoluindo" as melhores soluções ao longo de gerações.

Em resumo, quando hoje se fala em "IA", está-se quase sempre a falar de **Deep Learning** (Redes Neurais) e de outros tipos de **Machine Learning** (Supervisionado, Não Supervisionado e por Reforço).



Mini glossário

Educação mediática. Referenciais, iniciativas e projetos na União Europeia

Referenciais (documentos e estratégias)

Diretiva dos Serviços de Comunicação Social Audiovisual (AVMSD)

Esta é a principal peça legislativa. A sua versão revista (de 2018) **obriga legalmente** todos os Estados-Membros (incluindo Portugal) a "promover e tomar medidas para o desenvolvimento de competências de literacia mediática". Não é uma sugestão, é uma exigência.

Plano de Ação para a Educação Digital (Digital Education Action Plan)

Referencial estratégico da Comissão Europeia que define como a UE irá apoiar os Estados-Membros a adaptar os seus sistemas de educação e formação à era digital. Uma das suas prioridades é "reforçar as competências digitais e a literacia mediática", incluindo o combate à desinformação.

Grupo de Peritos em Literacia Mediática (MLEG)

A Comissão Europeia reúne regularmente este grupo de peritos (académicos, reguladores, empresas) para publicar relatórios e **linhas orientadoras** (referenciais) sobre as melhores práticas para ensinar e promover a literacia mediática na UE.

Projetos e Iniciativas

Programa Europa Criativa (Creative Europe)

Este é o principal instrumento de financiamento da UE para a cultura e os media. Tem uma vertente específica ("Media Literacy for All") que **financia projetos transnacionais** de literacia mediática, apoiando ONGs, universidades e meios de comunicação que queiram lançar iniciativas.

European Media Literacy Week

Uma "ação" de sensibilização pan-europeia. É uma campanha anual para destacar projetos locais e nacionais, incentivando a partilha de boas práticas entre os países.

Observatório Europeu dos Media Digitais (EDMO)

Esta é uma ação crucial. O EDMO é uma rede independente de verificadores de factos, académicos e jornalistas. Para além de combater a desinformação, uma das suas missões centrais é **desenvolver e disseminar materiais de literacia mediática** ao público.

O *hub* do EDMO para Portugal e Espanha chama-se **Iberifier**, que junta universidades e verificadores de factos (como o Polígrafo) para estudar a desinformação na Península Ibérica e promover a literacia.



Mini glossário

Educação mediática. Referenciais, iniciativas e projetos em Portugal

Referenciais

[Referencial de Educação para os Media \(para o Ensino Básico e Secundário\)](#)

Este é o documento "referencial" mais importante em Portugal. Foi criado pela **Direção-Geral da Educação (DGE)** em parceria com o **ERC (Entidade Reguladora para a Comunicação Social)** e outros peritos.

Iniciativa Nacional Competências Digitais e.2030 (Portugal Digital)

Esta é a estratégia nacional guarda-chuva para a transição digital. A "literacia digital e mediática" é um dos seus pilares centrais, com metas específicas para formar a população (jovens, adultos, trabalhadores) nestas competências.

Mandato do ERC

A própria lei da **Entidade Reguladora para a Comunicação Social (ERC)** define como uma das suas atribuições "incentivar a auto-regulação e a co-regulação" e "promover a literacia mediática".

Projetos e iniciativas

Projeto SeguraNet

Uma iniciativa do Centro de Competência TIC da DGE, focada na **Segurança Digital**, mas que tem uma enorme sobreposição com a literacia mediática (ex: como identificar perfis falsos, *fake news*, e proteger a privacidade).

Público na Escola

Uma das "ações" mais conhecidas e duradouras, gerida pelo jornal *Público*. Envia jornalistas às escolas de todo o país para explicar como funciona o jornalismo, como se constrói uma notícia e como se distingue de opinião ou desinformação.

Grupo Informal sobre Literacia Mediática (GILM)

Uma rede que agrega dezenas de organizações em Portugal – desde universidades (como o CICANT na Lusófona), bibliotecas, o ERC, a LUSA, o Polígrafo, e ONGs – para partilhar recursos e organizar eventos.

Apoio da DGLAB

A **Direção-Geral do Livro, dos Arquivos e das Bibliotecas** promove a literacia mediática através da Rede Nacional de Bibliotecas Públicas, que organizam workshops locais para várias faixas etárias.

Rede de Bibliotecas Escolares (RBE)

A literacia dos média faz parte do referencial “Aprender com a Biblioteca Escolar”.

Em 2024 a RBE lançou o programa ProLiteracias focado no desenvolvimento de competências em literacia mediática e informacional.



Mini glossário

Ad delivery — mecanismo algorítmico que decide a quem mostrar o anúncio dentro do público escolhido, otimizando objetivos do anunciante (pode introduzir vieses).

Astroturfing — é a criação artificial de uma aparência de apoio popular (“grassroots”) a uma causa, produto ou posição política. Organizações, campanhas ou empresas disfarçam-se de cidadãos comuns ou grupos locais para influenciar opinião pública, media e decisores, deturpando o consenso público, pressiona jornalistas e regulações, e pode contaminar consultas públicas ou eleições.

Este fenómeno implica financiamento e coordenação ocultos por parte de partidos, empresas, lobbies ou PACs. São criadas contas/organizações de fachada (associações “independentes” que não o são) com amplificação coordenada (bots, contas descartáveis, *influencers* pagos, *microtargeting*) e conteúdo “orgânico” fabricado: abaixo-assinados, comentários, cartas ao diretor, eventos.

Sinais de alerta para detetar o fenómeno: transparência fraca sobre quem financia e quem decide; existência de padrões de publicação sincronizados e repetição de mensagens idênticas e de “grupos locais” com domínios recentes, moradas vagas e contactos opacos com ligações indiretas a agências de comunicação.

Biblioteca de anúncios — repositório público (parcial) de anúncios, com campos como público, alcance, gasto.

Bot — conta ou software que publica/partilha conteúdo de forma automática ou semi-automática.

CIB (Comportamento Inautêntico Coordenado) — Rede de contas que age de forma concertada para manipular a conversa pública.

Conteúdo sintético (AIGC) — Texto, imagem, áudio ou vídeo gerado ou alterado por IA (inclui deepfakes).

Deepfake — Vídeo/áudio manipulado por IA para imitar pessoas reais de forma convincente.

"Dog whistle" (ou "apito de cão") — é um termo para uma mensagem codificada, muitas vezes política, que parece ter um significado inócuo para a maioria das pessoas, mas transmite uma mensagem específica a um público-alvo. É uma forma de comunicação que usa símbolos e linguagens que só são reconhecidos por um grupo específico, como extremistas de direita ou supremacistas brancos.

MAU (Monthly Active Users) — utilizadores ativos mensais.

Microdirecionamento (microtargeting) — entrega de mensagens/ anúncios personalizados a segmentos muito granulares usando dados comportamentais, demográficos ou psicográficos.

Proxy (indicador indireto) — métrica que não mede o fenómeno diretamente, mas ajuda a estimar a sua escala.

Robocalls — são chamadas telefónicas **automáticas** feitas por sistemas que discam números e reproduzem **mensagens gravadas** (ou voz sintetizada), em massa. Servem para marketing, avisos (ex.: companhias aéreas, escolas), cobranças ou propaganda política.

