

# Misinformation and disinformation in London: risks and responses

GLA City Intelligence rapid evidence review – March 2026

## 1. Introduction and key findings

This note summarises recent evidence on misinformation and disinformation, focusing on risks and potential responses that are relevant to London. It treats these issues as challenges to how information is understood and used. It does not cover related issues such as hate speech, harassment or other forms of online abuse.

The note draws on a rapid, high-level review of research and policy literature, alongside London-specific analysis. It is informed by expert input and stakeholder interviews with London-based organisations and practitioners.<sup>1</sup>

It addresses three main questions:

- What are misinformation and disinformation?
- Why do they matter for London and what risks do they pose?
- What actions appear most likely to reduce harms or improve resilience?

### Key findings:

- Misinformation and disinformation are now part of the UK's chronic risk landscape.<sup>2</sup> They pose risks to marginalised groups, democratic functioning, the economy and national security. London is particularly exposed given its global visibility, diversity and political prominence.
- Evidence shows that mis/disinformation can reduce trust in institutions, contribute to tensions between communities, and place pressures on public health and frontline services. Misleading narratives can also shape international perceptions of the capital, with potential economic impacts (eg, on investment and tourism).
- The evidence strongly suggests there is no single solution to address mis/disinformation challenges. It points to several practical areas where city and local actors can act through a multi-layered response. These include platform design, public communication, trusted messengers, media and information literacy and wider work to strengthen community resilience.
- The evidence base on impacts and specific interventions remains underdeveloped. Robust evaluation remains limited, and many approaches have not yet been tested systematically at city level or across local delivery settings.

---

<sup>1</sup> Particular thanks to Ika Trijsburg (Australian National University) and Paul Costello (German Marshall Fund Cities), and to others who provided comments on an earlier draft. Any errors remain the author's own. The analysis reflects available evidence and may be updated as further evidence emerges.

<sup>2</sup> Chronic risks are longer-term challenges that erode the UK economy, community, way of life and/or national security (HM Government, 2025). See the [glossary](#) for a summary of key definitions.

## 2. Framing the issue

UK sources generally distinguish between misinformation and disinformation in the following way. **Disinformation** refers to the deliberate creation and sharing of false or manipulated information, intended to mislead or cause harm. **Misinformation** refers to the unintentional spread of false or misleading information – for example rumours shared during fast-moving events (HM Government, 2025).

Sometimes, even information that is factual or partly true can have effects similar to misinformation or disinformation. For example, statistics or real events may be deliberately presented selectively or taken out of context to support misleading conclusions. This is often described as **malinformation** (GCS, 2025).

These distinctions are particularly important in contexts where they imply different types of response. For example, in national security matters or where deliberate influence campaigns or foreign interference are involved. As Purnat and Clark (2025) note:

‘Labelling something as misinformation is a value judgment on its accuracy. Disinformation is a value judgment on intent for distribution of inaccurate content. Each requires a different solution.’

In practice, especially for local actors, these definitions are often used together. This reflects how difficult it can be to determine intent, and that different types of misleading information can have similar effects. For example, content may begin as deliberate disinformation and later spread unintentionally as misinformation (POST, 2024). Ofcom research also notes that misinformation is partly a ‘subjective term’, reflecting differences in how people judge what counts as misleading (Ofcom, 2024b).<sup>3</sup>

Figure 1: Information types: fact and intent

False	<p><b>Misinformation</b></p> <p>False or misleading information presented as fact, not intended to cause harm</p>	<p><b>Disinformation</b></p> <p>Deliberately created and shared to mislead, harm or manipulate</p>
Fact	<p><b>Genuine information</b></p> <p>Factual information intended to inform</p>	<p><b>Malinformation</b></p> <p>Information based on fact, used out of context to mislead or cause harm</p>
		Intent to harm →

A simple framework for understanding misinformation, disinformation, malinformation and genuine information  
 Source: GLA Economics, adapted from Council of Europe (2017) and Deakin University Library • Created with Datawrapper

<sup>3</sup> In the rest of this note, ‘mis/disinformation’ is used as a shorthand for misleading or harmful information.

False or misleading information is not new. However, the current information environment has increased its speed, scale and reach. This reflects several developments, including:

- changes in how people access news and information – including growth in private messaging, social media and gaming platforms – alongside declining trust in news overall and rising risks of information overload<sup>4</sup>
- the role of algorithms in distributing online content, and monetisation models that reward engagement over accuracy
- attempts by domestic and foreign actors to influence or manipulate public opinion through misleading or deceptive information, often strategic and long-term, aimed at sowing confusion (Rycroft, 2026)
- the emergence of generative AI tools that lower the cost and level of skill required to produce persuasive content at scale, while making it harder for audiences to distinguish between authentic and altered material.

In this context, OECD analysis (2024) argues that policy should focus not only on individual pieces of content, but on strengthening the wider integrity of the information environment. This can be done through more transparent information systems, stronger societal resilience and more effective governance.

‘Dealing with hostile state online interference should be a far higher priority for government. This requires a better understanding of the problem and better coordination to deal with it.’

– The Rycroft Review, 2026

## Prevalence

It is difficult to measure the prevalence of mis/disinformation. Most surveys capture self-reported exposure to content that people believe may be false or misleading, rather than verifying whether the content is objectively false.

Recent Ofcom research gives some indication of scale. In a 2024 survey:

- 43% of London respondents said they had encountered misinformation in the previous four weeks – a figure in line with the UK average<sup>5</sup>
- 90% of all UK respondents said they were concerned about the societal impact of misinformation; 63% were concerned about its impact on them individually<sup>6</sup>
- respondents were most likely to report exposure online, with UK politics the topic most commonly associated with misinformation.

---

<sup>4</sup> Across the UK, the share of the population using print media as a source of news fell from 59% in 2013 to 12% in 2025, while the share using social media increased from 20% to 33%. From 2015 to 2025, trust in ‘most news most of the time’ fell from 51% to 35% (Reuters Institute, 2025; UK n=2,076).

<sup>5</sup> 39% had encountered written misinformation or disinformation, and 24% fake or deceptive images/videos. Source: Ofcom, 2024b (n=3667, London n=500).

<sup>6</sup> Ofcom, 2024b (n=1,580).

Wider Ofcom research suggests that while a majority of adult social media users report feeling confident judging whether online content was true or false, this confidence does not always match their response in objective skills tests (Ofcom, 2024c).

Qualitative research also suggests that exposure and impact vary by context. Anyone can be misled by mis/disinformation in certain circumstances.<sup>7</sup>

However, some groups – including younger adults, ethnic minority groups and people whose first language is not English – are disproportionately affected and may face more hostile information environments.<sup>8</sup> People in these groups may encounter misleading content more frequently, report greater concern about its effects or show stronger interest in building resilience (Ofcom, 2024a). Some studies also find lower accuracy in identifying false statements, though this likely reflects patterns of targeting and unequal access to trusted information sources rather than inherent vulnerability.

## Evidence gaps

The wider literature also highlights important evidence gaps.

According to POST (2024), there is ongoing disagreement about the overall prevalence of mis/disinformation. This is partly because offline circulation, passive exposure and private sharing (eg, through messaging services such as WhatsApp) are difficult to capture. Ofcom's (2025) literature review similarly concludes that much of the evidence on interventions is short term, frequently US-based and not always representative of real-world settings.

There are also limitations in understanding impacts. There is strong evidence that exposure to mis/disinformation can influence attitudes and behaviour in some contexts (eg, Faculty of Public Health, 2024). However, causal effects are often difficult to isolate, because behaviour change usually reflects multiple factors (POST, 2024).<sup>9</sup> Importantly, harms may also arise independently of behaviour change – for example, where individuals or groups are directly targeted by misleading information.<sup>10</sup>

And there are uncertainties about the effects of generative AI on mis/disinformation, particularly given the pace of development and evolving patterns of use (eg, Park and Nan, 2025). It may increase the volume of misleading content or undermine some interventions. At the same time, some early analysis suggests it could support delivery of interventions or expose users to more moderate or evidence-aligned information than some social media environments.<sup>11</sup>

---

<sup>7</sup> For example, research suggests that online misinformation appears to have influenced Covid-19 vaccination attitudes even among some London healthcare workers (UK Health Security Agency, 2026).

<sup>8</sup> These groups are all relatively highly represented in London. This suggests that the issues identified in national research may be particularly relevant in the London context.

<sup>9</sup> Rather than new information alone. Other factors affecting changes in behaviour include pre-existing beliefs, cultural/religious values, education, location, etc.

<sup>10</sup> Harms to targeted individuals or groups (eg, victims of hostile false narratives) can be experienced regardless of whether other people act on the mis/disinformation.

<sup>11</sup> See: Financial Times (2026) [Social media is populist and polarising; AI may be the opposite](#)

### 3. Why it matters for London

Wider UK evidence suggests that misinformation and disinformation can reduce trust in institutions, increase public anxiety and vaccine hesitancy, weaken crisis responses and add to tensions between communities.

Government risk assessments now treat mis/disinformation as part of the UK's chronic risk landscape, with potential harms for marginalised groups, the economy and national security (HM Government, 2025). The World Economic Forum's latest assessment ranked mis/disinformation as the second most severe short-term global risk (WEF, 2026).

#### Types of misleading narratives in London

London faces distinct risks within this wider context. Its global visibility, diversity and political and economic symbolism mean local incidents or narratives can be quickly amplified or reframed online – often reaching international audiences.

Misleading narratives tend to cluster around recurring themes. Analysis for the GLA identifies three main topics of online mis/disinformation:

- Claims that London is unsafe or in decline. These often draw on real offences (eg, sexual assault), but use selective reporting or decontextualised crime data to suggest wider patterns of decline. Related narratives focus on risks to women and girls, using emotive framing to target migrants or minority groups and promote more extreme views.
- Narratives relating to migration, population change and cultural identity. These includes claims about 'no-go' areas, 'Islamisation', demographic replacement or preferential treatment for particular groups.
- Narratives about policing and protest activity. These often intensify around specific events and include claims of 'two-tier' policing (the idea that police treat some groups differently to others).

Other recurring themes include antisemitic conspiracies, anti-LGBTQ+ and misogynistic narratives, and mis/disinformation relating to vaccines and public health, as well as climate, transport and urban planning policy.

#### How mis/disinformation operate in London

GLA research and interviews highlight several key features of these narratives and how they operate and spread in London.

**London's information environment is evolving quickly.** Social media analysis for the GLA indicates that online mis/disinformation relating to London has increased rapidly in recent years. Analysis of Reddit and other platforms also suggests a marked increase in narratives describing London as 'dangerous' or 'lawless' – a pattern not seen for other UK cities. Research by Dr Mark J. Hill (King's College London) found

posts containing these narratives increased from under 900 in 2008 to over 258,000 in 2024, alongside a rise in accounts assessed as 'strongly suspicious'.<sup>12</sup>

This suggests these narratives are not isolated spikes, but part of a broader and more sustained trend in volume and persistence.

Several factors contribute to this trend. These include a generally more permissive and less regulated platform environment; the scale and speed of AI-enabled content generation; hostile or state-linked activity; and London's continued prominence in debates on social cohesion, governance and demographic change. It is not clear from public evidence how far these dynamics are driven by foreign interference rather than domestic or organic activity (Rycroft, 2026). Declining trust in institutions and media also makes it easier for misleading narratives to gain traction.

There are, moreover, signs that mis/disinformation and extremist networks are becoming more connected with each other. Narratives can circulate across groups such as conspiracist, anti-vaccine and far-right communities.

**Events can also trigger rapid increases in false or misleading narratives.** Violent incidents, court judgments, geopolitical developments and contentious local issues can lead to spikes in mis/disinformation. Existing policy work points to the need for stronger local capability during periods of high tension and conflict (eg, Khan, 2024).

Interviewees described major events as accelerants, with false or misleading claims sometimes moving quickly into rumour, protest mobilisation or targeted hostility towards specific groups or communities. Research for the GLA shows that known disinformation amplifiers and hostile state-linked actors can selectively frame and amplify coverage of high-profile incidents. This can increase polarisation, encourage disorder and reinforce anti-London sentiment.

**These dynamics are reinforced by the fact that harmful narratives often draw strength from genuine grievances** – even where the framing is distorted. This matters because factual correction alone may not address the deeper reasons why such narratives resonate.

Interviewees described misleading claims that refer to real incidents or statistics but present them selectively or without context. Wider research shows that misleading narratives often develop within broader conflicts about trust and legitimacy. For example, a study of online narratives around low traffic neighbourhoods found that misinformation interacted with concerns about consultation and fairness, rather than emerging independently of them (Demos, 2024).

**The impacts are increasingly operational rather than purely theoretical.** Evidence suggests that mis/disinformation can place direct pressure on institutions and services. Interviewees reported local effects including harassment of staff, reputational targeting, increased security costs and the removal of public-facing information in response to

---

<sup>12</sup> See: Evening Standard (2026) ['London is not what it was': How social media accounts are changing the narrative of the capital.](#)

threats. These pressures affect community cohesion, safeguarding and public health, as well as operational capacity and budgets.<sup>13</sup>

### International perception and economic confidence

Mis/disinformation also create risks for London's international reputation and economy.

As misleading narratives about the city circulate and extend across global online networks, stakeholders involved in promoting London internationally report encountering them in their day-to-day work. Concerns about safety and stability are now a recurring feature of some international engagement, often requiring clarification and reassurance in discussions with investors and other contacts.

'Last year's narrative about rising crime in London showed how false information, largely circulating on social media, can shape perceptions of the city. During that period, we saw an increase in safety related enquiries from our international audiences.'

Laura Citron OBE, Chief Executive Officer, London & Partners

The impact on final decisions is difficult to measure. However, there is a growing risk that mis/disinformation undermine perceptions of London as a place to invest, visit, study and do business.

---

<sup>13</sup> A specific example is recent concerns and operational responses around a 'London school wars' meme. See: London Centric (2026) [How the London school wars meme spread from TikTok to mass panic – without a single reported police incident](#)

## 4. What actions appear most likely to reduce harms or improve resilience?

### National policy context

Central government approaches to countering misinformation and disinformation are mainly legislative and focused on monitoring disinformation trends (POST, 2024).

The main UK regulatory framework is the Online Safety Act 2023, with Ofcom responsible for oversight. National policy also places weight on media literacy and public resilience. While national regulation shapes the wider environment, many of the levers most relevant to London sit at city, borough and community levels.

### Potential interventions at city level

The evidence on responses to mis/disinformation at the city level is mixed. Some approaches are supported by systematic reviews or repeated findings across studies, while others are mainly proposed in policy guidance or frameworks. In general, there is more evidence around individual initiatives and short-term effects than on long-term changes in behaviour or wider system-level outcomes. Many interventions remain under-evaluated in the UK, particularly at the city level (eg, Ofcom, 2025).

Across the literature, one of the clearest implications for London is that no single response is likely to be enough. Effectiveness depends on how different approaches are combined. Reviews and policy frameworks emphasise multi-layered strategies that bring together different approaches (OECD, 2024; Trijsburg et al., 2024). Research also highlights that mis/disinformation is unlikely to be ‘solved’. Instead, it represents an ongoing challenge that must be managed and mitigated (POST, 2024).

‘There is no single way to address disinformation, and efforts need to include a combination of ongoing responses subject to frequent reflection and adaptation.’  
– *Disinformation in the City: Response Playbook* (Trijsburg et al., 2024)








The sections below review broad intervention types and their role within this layered response.

These interventions operate at different levels. Some shape the wider information environment, while others strengthen institutional responses, community communication or individual capability (Table 1). Assessments of evidence strength are indicative and reflect how interventions have been evaluated, which is typically in isolation. This limits understanding of how different approaches interact in practice and how effective they are when combined.<sup>14</sup>

---

<sup>14</sup> As noted, interventions are likely to be most effective when used together. For example, platform conditions may limit the spread of harmful content, institutions can respond quickly to emerging narratives and trusted messengers can help information reach audiences who are less likely to engage with official channels.

Table 1: Responses to misinformation and disinformation relevant at city level

Type		Intervention	Evidence strength
Information environment		Platform design and regulation	Strong evidence that platform conditions shape information spread
Institutional response		Public communication and cross-agency coordination	Strong for institutional relevance; limited causal evidence
Institutional response		Fact-checking and correction	Moderate evidence in specific settings
Community communication		Trusted local messengers	Promising but moderate evidence
Individual capability		Media and information literacy	Moderate to strong evidence of short-term effects
Individual capability		Prebunking and inoculation	Moderate to strong evidence of short-term effects
Social infrastructure		Social infrastructure and resilience	Promising but indirect evidence

*Evidence strength assessments are indicative. Evidence often assesses interventions in isolation, which may limit understanding of how approaches combine in practice. The table focuses on interventions relevant to city-level actors; platform design and regulation are included though largely sit at national or international level.*

Source: GLA City Intelligence desktop review of research and policy literature. • Created with Datawrapper

## Platform design and regulation

### What is it?

This refers to upstream platform-level measures that influence how information is distributed and assessed online. These include recommender systems, transparency and labelling measures, friction or sharing limits, content moderation practices, and wider regulatory frameworks.

### What does the evidence suggest?

A substantial body of research and policy analysis shows that platform design influences how misleading information is spread and how visible it becomes (Ofcom, 2025; CCDH, 2022). Design features can amplify or dampen content, shape how users interpret and evaluate information, and introduce friction that slows its circulation.

Exposure to even a small number of ‘hyperpartisan’ influencer accounts can have a significant effect on attitudes and behaviour (Rathje et al., 2025). OECD (2024)

highlights transparency, accountability and plurality of information sources as ways to improve the integrity of the information environment. Ofcom (2025) also considers how design and interface features can help users assess the credibility of online information.

More locally focused guidance also recognises the importance of the wider information environment. For example, the *Disinformation in the City: Response Playbook* and the UK RESIST counter-disinformation toolkit highlight monitoring online narratives and understanding platform dynamics as part of preparedness and response planning (Trijsburg et al., 2024; Government Communication Service, 2025).

For London, these levers are largely indirect. Most platform governance sits at national or international level, through legislation such as the Online Safety Act and regulatory oversight by Ofcom. Cities can still shape outcomes by engaging with regulators, working collectively and participating in international networks. In some cases, they may also influence platform behaviour through local policy or negotiation.<sup>15</sup> The main role for city government is therefore likely to involve monitoring emerging narratives, sharing evidence and contributing to regulatory debates.

These wider platform conditions also affect the effectiveness of other interventions and the context in which local responses take place.

### **Costs and risks**

Platform interventions often involve trade-offs between transparency, freedom of expression and enforcement. For city governments, the main limitation is that these levers are outside direct control. This means they must rely on national regulation, platform cooperation and broader information resilience measures.

### **Evidence strength**

Strong evidence that platform conditions influence how misleading information is spread; more limited evidence on which specific design interventions are most effective in practice, and on the role that city governments can play.

## **Prepared public communication and cross-agency coordination**

### **What is it?**

This refers to how public bodies prepare for, identify and respond to misleading narratives in a coordinated way. In practice, this can include agreed communications protocols, shared monitoring, clearer roles across agencies, and arrangements for coordinated messaging – particularly during fast-moving incidents.

### **What does the evidence suggest?**

This approach is strongly supported in policy and governance guidance. The OECD (2024), for example, sees public communication as a central part of the response to mis/disinformation. It highlights nine key principles, including institutionalisation, professional capability, transparency, prevention and strategic foresight. Playbook

---

<sup>15</sup> For example, Barcelona's engagement with Airbnb on short-term rental listings and data sharing.

guidance suggests that, alongside developing communications capacity, cities should play an important role in monitoring rhetoric and sentiment (Trijsburg et al., 2024).

For London, there is a strong case for improved public communication in policy and practical terms. However, the evidence reviewed largely suggests that this should form part of good practice, rather than demonstrating through strong evaluation that it consistently reduces harm. It also indicates that rapid institutional messaging can support other measures, including fact-checking efforts, and provide a reference point for trusted local messengers when they interact with communities.

### **Potential costs and risks**

This approach requires sustained capacity and coordination. It depends on trusted institutional relationships and can be difficult to maintain outside crisis periods. There is a risk that rapid official responses are seen as defensive, politicised or overly centralised, especially where trust is weak. Community listening can help improve public communication by guiding how messages are tailored (Ishizumi et al., 2024).

### **Evidence strength**

Strong for institutional relevance and practical plausibility; limited direct causal evidence on impacts.

### **Fact-checking and correction**

#### **What is it?**

Fact-checking and correction involve identifying false or misleading claims and responding with accurate information, clarification or rebuttal. This may be done by independent fact-checkers, media organisations, public bodies and platforms.

#### **What does the evidence suggest?**

Fact-checking remains important. Evidence shows that it can reduce belief in false claims when people see a correction (Ofcom, 2025). This means it has a stronger direct evidence base than some more structural approaches.

However, timing and reach are major constraints. Some evidence suggests that corrections often arrive after false claims have already spread widely, meaning they may not reach those most influenced by them (eg, Purnat and Clark, 2025). There is also disagreement about the most effective types of fact-check, while the OECD (2024) highlights the challenging environment in which fact-checkers operate. In a 2023 survey of 41 of Europe's leading fact-checking organisations cited by OECD, 90% reported some form of harassment, most of it online.

So, in some settings, fact-checking is supported by direct evidence. Yet the literature is clear that it is not enough on its own and is more effective when combined with other approaches.

#### **Costs and risks**

Fact-checking can be resource-intensive, reactive and difficult to sustain at pace. Its effectiveness drops where audience trust is low or where corrections circulate less widely than the original content. Generative AI may be used in ways that increase the

volume of mis/disinformation (though it may also support faster detection and verification). In some cases, prominent corrections can also unintentionally amplify a false claim or be undermined by perceived or actual politicisation.

### **Evidence strength**

Moderate evidence in specific settings; limited for broader system-level effects.

## **Community-based communication and trusted local messengers**

### **What is it?**

This approach involves working through people and organisations that local communities already trust, rather than relying only on formal institutional channels. These may include health and care workers, faith leaders, community organisations, schools, youth workers and other local networks.

### **What does the evidence suggest?**

This is one of the most relevant areas for London, particularly where mis/disinformation intersects with areas such as public health or social cohesion.

The evidence here is a mix of public health research, institutional frameworks, and practice-based reasoning. Reviews of vaccine and public health messaging suggest that community-based and tailored approaches can improve the reach and credibility of information. This is especially important where trust in official channels is low.<sup>16</sup>

Research in *The Lancet* argues for upstream, community-centred responses, rather than relying on rebuttal after false claims spread (Ishizumi et al., 2024). The Covid-19 response in London also shows the varied role trusted local intermediaries can play – sharing information through existing networks and feeding local concerns back to public authorities (ADPH London, 2023). More broadly, the OECD (2022) also places strong emphasis on partnering with non-governmental messengers in a transparent and open way. It highlights how digital tools can be used strategically to ‘encourage more responsive and inclusive communication’.

For London, this supports the case for working through local organisations, practitioners and community networks. However, the formal evidence base remains stronger on plausibility and reach than on quantified long-term impact.

### **Costs and risks**

Trusted-messenger approaches are highly context-specific. They depend on existing relationships, credibility, and local capacity. Delivery models vary and often rely on short-term funding or volunteers. This can make sustained engagement difficult, especially where messengers sit outside formal structures or lack stable resources or training. Effects are likely to vary across groups and settings.

### **Evidence strength**

Strong for practical relevance; promising but moderate for demonstrated impact.

---

<sup>16</sup> For example, some evidence suggests that people from certain ethnic minority groups are more likely to trust faith or community leaders (UK Health Security Agency, 2026).

## Media and information literacy

### What is it?

Media and information literacy aims to improve people's ability to assess sources, recognise misleading content, understand how information is produced and shared and make more informed judgements about what they encounter online and offline.

### What does the evidence suggest?

Media and information literacy is supported by a large and growing body of policy work and experimental evidence.

OECD (2024) presents it as part of a wider effort to strengthen information integrity across age groups. Ofcom's review finds that literacy-style interventions can improve awareness, confidence and the ability to recognise misleading content in the short term. Some research from the Covid-19 pandemic also found that higher levels of health, numerical and media literacy were associated with lower susceptibility to mis/disinformation (Ofcom, 2025). However, findings vary across studies, and evidence for longer-term behavioural change is weaker (Carnegie, 2025).

Schools are identified as one of the most effective settings for building these skills at scale – especially where media literacy is embedded across subjects and reinforced over time (eg, Department for Education, 2025; House of Lords, 2025). While work is underway, a recent House of Lords report (2025) notes that media literacy is not yet well embedded in the national curriculum, and provision varies between areas.

There is some programme-level evidence from the UK. The National Literacy Trust's (2024) evaluation of NewsWise in Birmingham reported improvements in pupils' confidence and fact-checking behaviours. The share correctly identifying two of three news items as real or fake increased from 33% before the project to 46% after it. These findings are useful, but they come from a local evaluation with a specific age group and do not demonstrate population-level effects over time.<sup>17</sup>

There is also a role for delivery through other local settings. Local community organisations (eg, libraries, youth clubs) can support access to media literacy for groups less likely to be reached through formal education. Evidence on these place-based approaches is still emerging<sup>18</sup> but they may improve reach and relevance.

### Costs and risks

Effects may be modest, uneven or short-lived. Levels of news literacy training are currently low.<sup>19</sup> And delivery at scale requires trained staff (eg, teachers) and sustained investment. It is also important to assess which forms of media literacy training remain effective over time (Carnegie, 2025) – particularly as generative AI makes misleading content easier to produce and harder to identify using traditional cues.

---

<sup>17</sup> Schools may also provide a route to engage parents and carers, extending reach beyond pupils, although evidence on these wider effects is limited.

<sup>18</sup> Ofcom is, for example, developing a 'place-based' model of media literacy as part of its three-year media literacy strategy (2024–2027). See: Ofcom (2024) [A Positive Vision for Media Literacy](#).

<sup>19</sup> A survey for the Reuters Institute found that only 16% of under 35s in the UK had received news literacy training, lower than in many other countries (Reuters Institute, 2025; also House of Lords, 2025).

Programmes may be harder to deliver effectively outside structured settings, especially for adults or groups embedded in polarised information environments. Some studies also note that promoting scepticism towards media can increase distrust of legitimate sources if not carefully designed, although the evidence on this point is mixed.

### **Evidence strength**

Moderate to strong for short-term knowledge and confidence gains; limited to moderate for sustained behavioural change at scale.

## **Prebunking and inoculation**

### **What is it?**

Prebunking and inoculation aim to build resistance to misleading narratives before people come across them. This often involves short interventions showing common techniques used in mis/disinformation – for example, emotional manipulation, scapegoating or false expertise – to make people more alert and resistant to them later.

### **What does the evidence suggest?**

In the academic literature, prebunking and inoculation are supported by behavioural approaches. However, the evidence needs to be framed carefully. Experimental and online studies suggest that prebunking can improve resistance to common misinformation techniques in the short term. Roozenbeek et al. (2022) found that psychological inoculation interventions improved resilience against online misinformation across several large-scale tests.

At the same time, much of this evidence is short term, often experimental, and frequently generated outside the UK. This limits confidence about long-term effectiveness in real-world settings (Ofcom, 2025).

Overall, this approach has a stronger research base than many others. However, the evidence on how well it translates into everyday local practice remains limited. Its relevance may grow if generative AI increases misleading content or weakens common credibility cues.<sup>20</sup>

### **Costs and risks**

The main risk is over-claiming. Effects observed in experiments may weaken over time or vary in real-world settings. Prebunking may also be less effective where distrust is already entrenched or where local actors do not have the credibility needed to deliver it effectively.

### **Evidence strength**

Promising but limited – useful short-term effects are supported by research, but there is limited evidence on sustained impact at city level.

---

<sup>20</sup> AI may also support prebunking by helping to identify emerging narratives and tailor messages, although this remains experimental and not yet proven at scale.

## Social infrastructure and resilience

### What is it?

This refers to the wider social conditions that help communities absorb shocks, maintain trust, and resist divisive narratives. It includes local relationships, neighbourhood institutions, community spaces and the broader social capital that supports collective action and dialogue within communities.

### What does the evidence suggest?

There is a growing body of work linking resilience to harmful narratives to the strength of local relationships, institutions and wider social infrastructure. For example, Coutts and Díaz Velásquez (2025) argue that social infrastructure and social capital can help communities resist divisive narratives and disorder. Qualitative research for Ofcom (2025) highlights the potential value of safe spaces – such as workshops online or in person – where people can express a range of views without judgement and engage with different perspectives. Emerging evidence suggests that moderated cross-community discussions can support perspective-taking and mutual understanding under certain conditions.

For London, this perspective is relevant because it links information resilience to wider work on social cohesion, including factors such as trust levels.<sup>21</sup> However, while some evidence suggests that stronger communities may be more resilient to mis/disinformation, this link is often inferred from broader research on social cohesion rather than directly evaluated in this context. Evidence on longer-term impacts is limited.

### Costs and risks

The link between social infrastructure and mis/disinformation outcomes is plausible, but it is also likely to be diffuse and difficult to measure. Interventions in this area also tend to operate over long timeframes. Qualitative research further suggests that participation in moderated dialogue or workshop settings can be uneven and motivating some groups to engage can be challenging.

### Evidence strength

The evidence is promising but indirect.

---

<sup>21</sup> According to DCMS estimates for 2024/25, the proportion of adults in London (31%) who report that many people in their neighbourhood can be trusted is lower than in other regions (40% England-wide). Source: DCMS (2025) [Community Life Survey 2024/25: Neighbourhood and community](#).

## 5. Implications for London policy and practice

Overall, the evidence suggests that London should treat misinformation and disinformation as a cross-cutting resilience issue affecting its international reputation, public service delivery, democratic functioning and wider social cohesion.

The evidence does not show that any single intervention can address mis/disinformation challenges. It points to several practical areas where city and local actors can – and should – act. These areas include action across the information environment, public communication, trusted local networks and individual capability as part of a layered response.

Importantly, the evidence base is still underdeveloped. There is clear value in improved local monitoring, stronger evaluation of responses and more city-level evidence on what works over time. Existing evidence also tends to assess interventions in isolation, which limits understanding of how different approaches combine.

Developments in generative AI are also likely to change London's information environment, particularly given the pace of development and evolving patterns of use. This may affect the effectiveness of existing approaches and reinforces the need for ongoing review and adaptation.

## Short glossary

**Acute risks** – discrete events requiring an emergency response, which would have a substantial impact on the UK's safety, security and/or critical systems at a national level. Assessed in the government's [National Risk Register](#).

**Chronic risks** – longer-term challenges that erode the UK economy, community, way of life and/or national security. They can increase the likelihood and impact of acute risks. Identified and assessed in the government's [Chronic Risks Analysis](#).

**Misinformation** – false information shared without intent to cause harm or mislead, though UK bodies often note that intent is hard to observe.

**Disinformation** – false or misleading information created or shared deliberately to mislead or cause harm.

**Malinformation** – genuine information used to cause harm, for example through strategic leaking or harassment.

**Prebunking / inoculation** – efforts to prepare people for misleading techniques before they encounter them.

**Prevalence** – usually a measure of reported exposure, not the same as belief, sharing or behavioural effect.

**Upstream** – used to refer to actions taken early to prevent or reduce the spread of misleading information, rather than responding after it has spread.

## Selected bibliography

### UK parliamentary, regulator and government

DSIT, DCMS and DfE (2026) [A safe, informed digital nation: government's media literacy action plan \(2026–2029\)](#)

Department for Education (2025) [Curriculum and Assessment Review Final Report](#)

Government Communication Service (GCS) (2025) [RESIST 3: Building resilience to information threats](#)

HM Government (2025) [Chronic Risks Analysis](#)

House of Commons Foreign Affairs Committee (2026) [Disinformation Diplomacy: Oral evidence, HC 703, 13 January 2026](#)

House of Commons Science, Innovation and Technology Committee (2025) [Social media, misinformation and harmful algorithms](#)

House of Lords Communications and Digital Committee (2025) [Media literacy](#)

Khan, S. (2024) [The Khan Review: Threats to Social Cohesion and Democratic Resilience](#)

Ofcom (2024a) [Co-creating ways to navigate and mitigate against mis and disinformation](#)

Ofcom (2024b) [Mis and disinformation report](#)

Ofcom (2024c) [Adults' Media Use and Attitudes Report](#)

Ofcom (2025) [Misinformation and disinformation: literature review](#)

Parliamentary Office of Science and Technology (POST) (2024) [Disinformation: sources, spread and impact](#)

Rycroft (2026) [The Rycroft Review: Report of the independent review into countering foreign financial influence and interference in UK politics](#)

UK Parliament (2023) [Online Safety Act 2023](#)

UK Health Security Agency (2026) [National Immunisation programme: Health Equity Audit 2025](#) (Appendix 4. Narrative review of the determinants of vaccination uptake)

### **Local government, cities and practice-focused sources**

ADPH London (2024) [Transformations in Community Collaboration: Lessons from COVID-19 Champions across London](#)

C40 Cities (2024) [Extreme Weather Disinformation Playbook](#)

Counter Terrorism Preparedness Network (2023) [Mis- and Disinformation: Extremism in the Digital Age](#)

Local Government Association (2026) [A guide to disinformation affecting local authorities and their communities](#)

Trijsburg, et al. (2024) [Disinformation in the City: Response Playbook](#)

### **Think tanks and civil society**

Coutts, et al. (2025) [Social infrastructure and social capital – the active elements of community resilience: A policy review for the Independent Commission on Neighbourhoods](#)

Demos (2024) [Driving Disinformation: Democratic deficits, disinformation and low traffic neighbourhoods – a portrait of policy failure](#)

Demos (2025) [Community Disorder: How do we prevent an information emergency?](#)

HOPE not hate (2023) [Plugged In but Disconnected](#)

HOPE not hate (2024) [Fear and Hope 2024](#)

National Literacy Trust (2024) [Evaluation of NewsWise in Birmingham](#)

Race Equality Foundation (2025) [Understanding the Racist Riots of 2024 and what should be done](#)

Reuters Institute (2025) [Reuters Institute Digital News Report 2025](#)

The Royal Society (2022) [The online information environment](#)

### **International / academic**

Carnegie Endowment for International Peace (2024) [Countering Disinformation Effectively: An Evidence-Based Policy Guide](#)

Council of Europe (2017) [Information Disorder: Toward an interdisciplinary framework for research and policy making](#)

Council of European Municipalities and Regions (CEMR) (2025) [Local truth, shared trust: local and regional responses to disinformation](#)

OECD (2020) [Governance responses to disinformation: How open government principles can inform policy options](#)

OECD (2022) [Good Practice Principles for Public Communication Responses to Mis- and Disinformation](#)

OECD (2024) [Facts not Fakes: Tackling Disinformation, Strengthening Information Integrity](#)

Park and Nan (2025) [Generative AI and misinformation: a scoping review of the role of generative AI in the generation, detection, mitigation, and impact of misinformation](#)

Purnat and Clark (2025) [Oversimplified efforts to counter health misinformation are missing the mark](#)

Rathje et al. (2025) [Unfollowing hyperpartisan social media influencers durably reduces out-party animosity](#)

Roozenbeek, et al. (2023) [Countering Misinformation: Evidence, Knowledge Gaps, and Implications of Current Interventions](#)

Van der Linden, et al. (2022) [Prebunking interventions based on the psychological theory of “inoculation” can reduce susceptibility to misinformation across cultures](#)

Borges do Nascimento et al. (2022) [Infodemics and health misinformation: a systematic review of reviews](#). World Health Organization

World Economic Forum (WEF) (2026) [Global Risks Report 2026](#)

### **Other sources (non-specific)**

Centre for Countering Digital Hate (CCDH) [Research on online harms and disinformation](#)

Full Fact [Full Fact's Research](#)

Neighbourly Lab [Research on social cohesion, polarisation and community trust](#)