

ChatGPT explains EU Platform Law

Daphne Keller, Aug 19 2024

I spent some time last week playing with this “[Trust & Safety Regulation expert](#)” version of ChatGPT. It is, as far as I know, very much an experiment -- no one is suggesting relying on it for legal advice. I asked it some very basic compliance questions, and a few gotchas. Its answers about both [US](#) and EU law were both enlightening and alarming. I think they tell us a lot about the laws themselves, and about the secondary materials that companies will rely on in complying with them.

The regulation-focused ChatGPT consistently told platforms to take down potentially or actually lawful expression posted by Internet users. It also made a number of mistakes, both in “gotchas” and more basic questions about laws like the DSA, GDPR, and DSM Copyright Directive Article 17. Often these are the same mistakes my Stanford students, who probably also train on some Internet summaries of the laws, make on exams. But they are also generally the answers that you would expect from a purely compliance-minded lawyer whose goal is to build the right processes, check the boxes, create the documentation, and not rock the boat. (I tried running these same questions by the normal version of ChatGPT and got [answers](#) that were broadly similar, but far less detailed and less focused on the operational details of compliance. [Graham Smith](#) tried pressing the regulation-focused ChatGPT on its “monitoring” recommendations, and it [backed down](#) quite a bit when confronted with case law.)

Legal experts see nuance or carefully negotiated protections for fundamental rights in these laws. ChatGPT is, perhaps, saying the quiet part out loud. It is holding up a mirror to its training materials and prompts -- meaning, presumably, the laws themselves and fairly standard secondary sources explaining them.

Here are the questions I asked it about EU law.

I. DSA Risk Mitigation

In ChatGPT's hands, important contested aspects of the DSA get resolved firmly on the side of overcompliance. I asked it a question about risk mitigation (meaning Article 34 and 35), but I deliberately used language from Commissioner Breton's most recent [letter](#) to X in referring to those obligations.

ChatGPT repeatedly told me to remove specific categories of potentially lawful but “harmful” content. That's not what the DSA was supposed to require. As one leading expert [explains](#),

enforcers “cannot tell providers what lawful expressions they must prohibit or suppress on their services[.]” During legislative drafting, the Commission’s position was that the law was only about removing (1) illegal content and (2) content the platform itself prohibited in its TOS. It was not a mechanism for EU lawmakers to identify and suppress new categories of expression and information. As platforms struggle to actually do the required “risk mitigation,” though, it is hard to avoid the conclusion that this really means removing harmful but lawful content.

ChatGPT’s answers also have some ambiguity as to whether obligations from the rest of the DSA (re dark patterns, minors, advertising, notice and action processes, transparency etc.) all get reincorporated and reconsidered under Articles 34 and 35. If they do, then the carefully drafted and negotiated prescriptive language in Articles 8-32 would be largely a dead letter, replaced by vague “risk” analysis. ChatGPT’s response here could be partly my prompt’s fault, since I used the “risk mitigation” language without specifying that I meant Articles 34 and 35.

PROMPT: I run a social media platform in Italy. What must I do to mitigate risks of detrimental effects on civic discourse and public security under the DSA?

RESPONSE:

To mitigate the risks of detrimental effects on civic discourse and public security under the **Digital Services Act (DSA)**, your social media platform in Italy must implement several key measures. The DSA requires platforms to **take proactive steps** to minimize the spread of **harmful content**, protect public security, and ensure a healthy online environment. Here’s what you should do:

1. Conduct Regular Risk Assessments

- **Systemic Risk Identification:** Regularly assess and identify systemic risks associated with your platform, particularly those that could impact civic discourse and public security. This includes the spread of disinformation, hate speech, and content that could incite violence or public unrest.
- **Risk Mitigation Plans:** Develop and implement risk mitigation strategies based on the findings of these assessments. This might involve **adjusting your content moderation policies**, improving algorithms, or enhancing user reporting systems.

2. Implement Robust Content Moderation

- **Proactive Monitoring:** Use advanced content moderation tools, including AI and machine learning, to **detect** and remove illegal content, such as hate speech, terrorist content, and disinformation, before it can cause harm.
- **Human Oversight:** Ensure that there is human oversight in the moderation process, particularly for complex or borderline cases where automated systems might struggle to make nuanced decisions.

- **Transparency in Moderation:** Provide clear and transparent guidelines on how content is moderated, including what types of content are not allowed and the rationale behind these decisions.

3. Enhance Transparency and Reporting

- **Transparency Reports:** Publish regular transparency reports detailing your content moderation efforts, including the number of content removals, types of content removed, and the reasons for removal. This increases accountability and public trust.
- **Public Disclosures:** Ensure that your platform's terms of service, community guidelines, and moderation practices are easily accessible and clearly explained to users.

4. Strengthen User Empowerment Tools

- **Reporting Mechanisms:** Provide users with easy-to-use tools to report **harmful** content. Ensure that these reports are reviewed promptly, and take appropriate action where necessary.
- **Appeals Process:** Establish a clear and accessible process for users to appeal content moderation decisions, **ensuring** fairness and transparency in how these decisions are handled.
- **Content Labels and Warnings:** Consider adding labels or warnings to content that might be **harmful** but does not meet the threshold for removal. This helps users make informed decisions about the content they engage with.

5. Coordinate with Authorities

- **Cooperation with Italian Authorities:** Maintain open lines of communication with Italian law enforcement and regulatory authorities, such as AGCOM (Autorità per le Garanzie nelle Comunicazioni), to ensure compliance with local laws and regulations.
- **Emergency Protocols:** Develop and implement protocols for rapidly addressing content that could pose a significant threat to public security, such as during elections, public health emergencies, or civil unrest.

6. Audit and Improve Algorithms

- **Algorithmic Accountability:** Regularly audit the algorithms used for content recommendation and dissemination to **ensure** they do not amplify **harmful** content or **contribute to the polarization** of civic discourse.
- **Bias Mitigation:** Implement measures to identify and mitigate any biases within your algorithms that could disproportionately impact certain groups or viewpoints.

7. User Education and Awareness

- **Digital Literacy Initiatives:** Invest in digital literacy programs to educate your users about recognizing disinformation, understanding the importance of civil discourse, and using your platform responsibly.

- **Community Engagement:** Foster a community culture that promotes respectful discourse and discourages harmful behavior. This can be supported through community guidelines, positive reinforcement, and user engagement strategies.

8. Compliance with Legal Obligations

- **Data Access and Transparency:** Be prepared to provide data access to regulators for monitoring compliance with the DSA. Ensure that your data management practices are robust and that you can supply the required information when needed.
- **Legal Response Protocols:** Develop protocols for responding to legal requests for content removal or information sharing in a timely and compliant manner.

Conclusion

To mitigate risks under the DSA, your platform **must actively manage** the spread of **harmful** content, engage transparently with users and regulators, and **ensure** that your content moderation practices protect both public security and the integrity of civic discourse. Regularly reviewing and updating your processes in response to emerging risks and regulatory changes is crucial for ongoing compliance and the safety of your user base.

A. Variant of the Risk Mitigation Question

In an earlier iteration of this question, I asked about what I “can” so instead of what I “must” do. Its answers were similar, with these additions:

PROMPT: I run a social media platform in Italy. How can I mitigate risks of detrimental effects on civic discourse and public security under the DSA?

EXCERPTS FROM RESPONSE:

- **Implement Robust Content Moderation:** Ensure your platform has a comprehensive system for **monitoring** and moderating **harmful** content, especially disinformation and illegal content that could harm public security or disrupt civic discourse. The DSA emphasizes the need for online platforms to remove illegal content promptly and to **prevent** the dissemination of harmful disinformation.
- ...
- **Compliance with Legal Orders:** Ensure your platform can quickly comply with legal orders to remove content that poses a threat to public security. The DSA mandates timely responses to such orders, particularly concerning disinformation and hate speech.

II. Copyright Article 17 Filters

It has been about three years since I actively engaged with this law, so I am more hesitant about my own take on the law. The European colleagues who I asked are all on vacation.

PROMPT: What do I need to do to comply with Copyright Directive (also known as Digital Single Market Directive) Article 17, can I implement the same procedures in France, Finland, and Germany?

RESPONSE:

Compliance with **Article 17** of the **Copyright Directive** (also known as the Digital Single Market Directive) requires platforms that host user-uploaded content to implement specific measures to **ensure** that copyrighted works are not shared without authorization. While Article 17 sets out harmonized obligations across the EU, the implementation of these obligations can vary slightly by country due to national transpositions. Here's what you need to do and considerations for applying the same procedures in **France, Finland, and Germany**:

Core Requirements Under Article 17

1. Content Filtering Obligations:

- **Preventive Measures:** Platforms must take effective measures, such as content recognition technologies, to **prevent the unauthorized sharing** of copyrighted content.
- **Licensing Agreements:** Platforms are encouraged to negotiate licensing agreements with rights holders to cover the use of their works. If such agreements are in place, users can upload content without restrictions.
- **Notice and Takedown Procedures:** Platforms must have robust systems in place to promptly remove or disable access to content when rights holders notify them of a copyright infringement.

2. User Rights Protection:

- **Complaint and Redress Mechanisms:** Users must have access to mechanisms that allow them to contest the removal of their content. This is especially important when the content is removed erroneously, such as in cases of fair use or parody.
- **Transparency and Reporting:** Platforms must provide users with information about their content moderation practices and decisions, especially when content is removed or access is disabled.

Country-Specific Considerations

While Article 17 is EU-wide, each member state may have slightly different approaches to enforcement and implementation, which may affect your compliance strategy.

1. France

- **Implementation:** France has fully implemented the Copyright Directive into its national law. The French authorities have been particularly proactive in requiring platforms to prevent the sharing of unauthorized content.
- **Key Considerations:**
 - **Strict Enforcement:** France may expect platforms to adopt more rigorous content filtering technologies to comply with Article 17, given its strong stance on protecting creators' rights.
 - **User Rights:** Ensure that your platform's complaint mechanisms are robust and clearly communicated, as France places significant emphasis on balancing copyright protection with user rights.

2. Finland

- **Implementation:** Finland has implemented the Directive in a manner consistent with broader EU guidelines, focusing on a balanced approach between protecting copyright and safeguarding user rights.
- **Key Considerations:**
 - **Balanced Approach:** Finland's implementation tends to focus on ensuring that measures taken are proportionate and that user rights are not unduly restricted.
 - **Licensing and Filtering:** Ensure your platform's content filtering mechanisms are in line with the principle of proportionality and that licensing agreements are clearly defined.

3. Germany

- **Implementation:** Germany has implemented the Directive with particular attention to the balance between copyright enforcement and freedom of expression. Germany introduced the concept of "de minimis" use, allowing for minor uses of copyrighted content without triggering a takedown.
- **Key Considerations:**
 - **De Minimis Use:** In Germany, minor and non-commercial uses of copyrighted works are permitted, which requires your platform to differentiate between significant and minor uses when filtering content.
 - **User Rights Emphasis:** Germany places a strong emphasis on user rights, requiring a detailed and fair appeals process for users whose content is removed.

Can You Implement the Same Procedures?

While the core principles of Article 17 are consistent across the EU, you will need to **tailor your approach** slightly to fit the specific legal nuances in France, Finland, and Germany:

- **Content Filtering:** Implement robust filtering technologies across all three countries but be particularly mindful of Germany's "de minimis" provisions, which might require more nuanced filtering practices.
- **Licensing:** Secure comprehensive licensing agreements with rights holders that are applicable across the EU. Ensure these agreements are flexible enough to accommodate country-specific requirements.
- **User Rights and Redress Mechanisms:** Develop a user rights and appeals process that complies with the stricter requirements in Germany and France, while ensuring transparency and fairness in Finland.

Conclusion

You can implement largely similar procedures across France, Finland, and Germany to comply with Article 17, but you need to account for each country's specific legal nuances. This involves making minor adjustments in how you apply content filtering, handle minor uses, and protect user rights to ensure full compliance across these jurisdictions.

III. Right to Be Forgotten

I published a detailed [article](#) in 2018 about this question and the resulting intersection of (1) the GDPR and (2) the EU's intermediary liability laws, then grounded in the eCommerce Directive and now in the DSA. I have not tracked developments closely since then, but I think I would know about anything major. Some things may have changed, including evolution of critical questions about the application of GDPR Art 17 ("Right to Be Forgotten") to hosting intermediaries like Facebook or TikTok.

PROMPT:

I run a social media company in Europe. A user alleges that a photo of her that a third party posted violates her data protection rights. What should I do?

RESPONSE:

When a user alleges that a photo posted by a third party on your social media platform violates their data protection rights, it's crucial to respond promptly and in accordance with the **General Data Protection Regulation (GDPR)**. Here's a step-by-step guide on how to handle the situation:

1. Acknowledge the Complaint

- **Prompt Response:** Acknowledge the user's complaint immediately, confirming that you have received it and will investigate the issue.

- **Explain the Process:** Inform the user of the steps you will take to address the complaint, including any necessary timeframes.

2. Assess the Complaint

- **Review the Content:** Examine the photo in question to determine whether it could potentially violate the user's data protection rights under GDPR.
- **Consider GDPR Criteria:**
 - **Personal Data:** Confirm whether the photo qualifies as personal data (i.e., it identifies the user directly or indirectly).
 - **Consent:** Determine whether the user has given consent for the photo to be posted. Under GDPR, consent must be freely given, specific, informed, and unambiguous.
 - **Legitimate Interest:** Consider if the third party had a legitimate interest in posting the photo and whether that interest outweighs the user's right to privacy.
 - **Special Categories of Data:** Check if the photo contains special categories of personal data (e.g., data revealing racial or ethnic origin, health data), which require stricter protections.

3. Take Action Based on Findings

- **If a Violation is Found:**
 - **Remove the Content:** If the photo violates the user's data protection rights, remove or restrict access to the photo as soon as possible.
 - **Notify the User:** Inform the user of the actions taken and their right to seek further remedies if they are not satisfied with your response.
 - **Notify the Third Party:** Inform the third party who posted the photo about the removal and the reason behind it. If applicable, explain the basis under GDPR for the removal.
- **If No Violation is Found:**
 - **Explain the Decision:** If you determine that the photo does not violate the user's rights, explain your reasoning to the user in a clear and transparent manner.
 - **Offer Redress Options:** Provide the user with information on how they can escalate their complaint, such as contacting the Data Protection Officer (DPO) or the relevant Data Protection Authority (DPA).

4. Document the Process

- **Record Keeping:** Document the entire process, including the complaint, your investigation, the decision made, and any actions taken. This is important for compliance and in case of future disputes.

5. Review and Update Policies

- **Compliance Check:** Ensure that your platform's policies and procedures align with GDPR, particularly regarding user rights, consent management, and content moderation.
- **User Education:** Consider providing users with resources or guidelines on how to protect their privacy and what to do if they believe their rights have been violated.

6. Cooperate with Authorities

- **Report if Necessary:** If the complaint involves a serious violation or the user escalates the issue, you may need to report it to the relevant Data Protection Authority. Be prepared to cooperate fully with any investigation.

Key Points to Remember:

- Under GDPR, users have the right to request the removal of content that violates their data protection rights, particularly if they have not consented to the use of their personal data.
- Your response should be timely, transparent, and in full compliance with GDPR obligations.

By following these steps, you can ensure that your social media platform respects the user's data protection rights and complies with European data protection laws.

A. Right to Be Forgotten and Health Info in a Photo

PROMPT:

the photo shows the person with an arm injury, their arm is in a cast. is that sensitive health information? how should it affect my decision?

RESPONSE:

Yes, a photo showing a person with an arm injury, particularly with their arm in a cast, can be considered **sensitive health information** under the **General Data Protection Regulation (GDPR)**. Here's why and how it should influence your decision:

1. Classification as Sensitive Data

- **Health Data Definition:** Under GDPR, health data is considered a special category of personal data that requires stricter protection. Health data includes information related to

the physical or mental health of an individual, which reveals information about their health status.

- **Photo of an Arm in a Cast:** A photo showing an arm in a cast directly relates to the individual's health status, as it indicates that the person has suffered an injury. This qualifies as sensitive health data.

2. Impact on Your Decision-Making Process

- **Stricter Requirements for Processing:** Since the photo is considered sensitive data, GDPR imposes stricter conditions for its lawful processing. Generally, you would need the explicit consent of the individual to process or display such data publicly.
- **Consent Review:** Determine whether the user gave explicit consent for the photo to be posted. If not, the processing of this data (i.e., displaying the photo) is likely unlawful under GDPR.
- **Legitimate Grounds:** Even if consent was not given, consider whether there are other legitimate grounds for processing this data, such as vital interests, legal obligations, or a legitimate interest that outweighs the user's privacy rights. However, these grounds are more difficult to justify for sensitive data.

3. Recommended Actions

- **Immediate Removal:** If the user did not provide explicit consent for the photo to be posted and no other legal basis applies, you should remove the photo promptly to comply with GDPR.
- **Notify the User:** Inform the user of the actions taken, explaining that the photo was removed because it constitutes sensitive health data and that GDPR requires explicit consent for such data to be processed.
- **Inform the Third Party:** Notify the third party who posted the photo about the removal and the legal reasons behind it, emphasizing GDPR's strict requirements for processing sensitive health data.

4. Documentation and Reporting

- **Document the Process:** Keep a detailed record of the complaint, your investigation, and the actions taken. This documentation is crucial in case the decision is later challenged.
- **Possible Reporting:** Depending on the circumstances, if the posting of the photo represents a significant data breach, you may need to report the incident to the relevant Data Protection Authority (DPA) within 72 hours.

Conclusion

Since the photo reveals health information, it is subject to stricter GDPR rules. If the user did not explicitly consent to the posting of the photo, you should remove it immediately. This action will

ensure that your platform remains compliant with GDPR's requirements for handling sensitive personal data.

B. Right to Be Forgotten and Public Interest

PROMPT:

The person who posted the photo says it depicts their town's former mayor in the act of entering the house where his mistress lives. It was published in order to let people know about his adultery. Should I put it back up?

RESPONSE:

Given the sensitive nature of the content and the legal implications, the decision to reinstate the photo should be carefully considered, particularly in light of GDPR and other relevant legal principles.

1. Assessing the Legal Grounds for Reposting

- **Personal Data and Public Interest:** The photo likely contains personal data since it identifies an individual (the former mayor) and relates to their private life. While public interest, particularly in cases involving public figures, can sometimes justify the publication of personal information, GDPR requires a careful balancing of the individual's privacy rights against the public's right to information.
- **GDPR Considerations:** Under GDPR, processing sensitive personal data (which can include data related to an individual's private life) requires a lawful basis. The public interest or freedom of expression may provide a lawful basis, but this depends on the context and the legal framework in the specific country.

2. Balancing Privacy and Freedom of Expression

- **Privacy Rights:** The former mayor's privacy rights are protected under GDPR, especially given that the image reveals information about his private life (his relationship with his alleged mistress).
- **Freedom of Expression:** On the other hand, if the former mayor is a public figure, particularly a political one, there may be a stronger argument for freedom of expression and public interest, especially if the information is relevant to his role or public duties.
- **Proportionality Test:** You need to consider whether the publication of the photo is proportionate to the public interest served. Does the public have a legitimate need to know this information? Is there a less intrusive way to achieve the same goal?

3. Jurisdiction-Specific Considerations

- **Defamation and Privacy Laws:** Beyond GDPR, you need to consider local defamation and privacy laws, which might further influence whether the image can be reposted. In some jurisdictions, the balance between privacy and public interest is weighted differently, and defamation laws might prevent the reposting of the image without clear evidence supporting the public interest claim.
- **Right to Be Forgotten:** The former mayor might have a right to request the removal of the photo under the “right to be forgotten,” especially if the content is deemed unnecessary or irrelevant for the public discourse.

4. Next Steps

- **Consult Legal Counsel:** Before deciding to repost the image, it’s advisable to consult with legal counsel who is familiar with privacy and media law in your jurisdiction. They can provide guidance on whether the photo’s publication is defensible under local laws.
- **Consider the Context:** If the former mayor is no longer a public figure or if the information is not of public interest, it is more likely that privacy rights will outweigh the justification for reposting the image.
- **Final Decision:** If you decide not to repost the image, clearly communicate the reasons to the person who posted it, focusing on the legal and ethical considerations. If you do decide to repost it, ensure that you have documented your decision-making process and the legal grounds supporting it.

Conclusion

Given the potential legal risks and ethical concerns, it’s important to weigh the privacy rights of the former mayor against the public interest. In most cases, unless there is a compelling public interest that justifies the invasion of privacy, the safer course of action would be not to repost the photo.