IS IT A PLATFORM? IS IT A SEARCH ENGINE? IT'S CHAT GPT! THE EUROPEAN LIABILITY REGIME FOR LARGE LANGUAGE MODELS

Beatriz Botero Arcila[*]

[*] Assistant Professor of Law, Sciences Po Law School, Paris and Faculty Associate, Berkman Klein Center for Internet and Society at Harvard University (beatriz.boteroarcila@sciencespo.fr). Special thanks to Rachel Griffin and Raphaële Xenidis for their feedback, to Jimena Escobar for editing assistance and to Eugene Volokh for his generous editing and feedback. I used ChatGPT to ask some questions and for help footnoting. All mistakes are mine.

Introduction

ChatGPT and other AI large language models (LLMs) raise many of the regulatory and ethical challenges familiar to AI and social media scholars: They have been found to confidently invent information and present it as fact.[1] They can be tricked into providing dangerous information even when they have been trained to not answer some of those questions[2]—such as giving advice on how to plan an attack or how to build a Molotov cocktail if asked through hypotheticals.[3] They can output detailed arguments very quickly, which may make the cost of producing disinformation very low (though some have argued that this risk is overblown because that cost is already very low).[4] Their ability to mimic a personalized conversation can be very persuasive, which creates important disinformation and fraud risks.[5] They reproduce various societal biases, because they are trained on data from the internet that embodies such biases, for example on issues related to gender and traditional work roles.[6] They have already started raising data protection and security concerns, as shown by a first leak of user data in late March 2023 and Italy's data protection agency's temporary ban of ChatGPT.[7] Thus, like other AI systems, LLMs risk sustaining or enhancing discrimination and perpetuating bias, and

---

[1] James Vincent, *OpenAI's New Chatbot Can Explain Code and Write Sitcom Scripts but Is Still Easily Tricked*, The Verge (Dec. 1, 2022), https://perma.cc/ASB4-G7C2; Jane Bambauer, *Negligent AI Speech: Some Thoughts About Duty*, 3 J. Free Speech L. 343 (2023).

[2] *See* OpenAI, *GPT-4 Technical Report*, ArXiv:2303.08774 (Mar. 21, 2023), https://perma.cc/6MTM-GWKS [hereinafter OpenAI, *Technical Report*].

[3] Zack Switten (@zswitten), Twitter (July 22, 2022), https://perma.cc/6Z58-WGG8.

[4] *Id.*; *see also* Arvind Narayanan & Sayash Kapoor, *The LLaMA Is out of the Bag. Should We Expect a Tidal Wave of Disinformation?*, Knight Institute, Algorithmic Amplification and Society (Mar. 6, 2023), https://perma.cc/G3V3-45TC.

[5] Narayanan & Kapoor, *supra* note 4.

[6] *See* Rory Gills, Brent Mittelstadt & Sandra Wachter, *ChatGPT—Friend or Foe?*, Oxford Internet Institute (Mar. 14, 2022), https://perma.cc/235H-F3Y4; Davey Alba, *OpenAI Chatbot Spits Out Biased Musings, Despite Guardrails*, Bloomberg (Dec. 8, 2022), https://perma.cc/4RXX-DXTB; OpenAI, *Technical Report, supra* note 2, at 42.

[7] *See* Shiona McCallum, *ChatGPT Banned in Italy over Privacy Concerns*, BBC (Apr. 2, 2023), https://perma.cc/QS2N-WNP7.

promoting the growth of corporate surveillance, while being technically and legally opaque.[8] Like social media, LLMs pose risks associated with the production and dissemination of information online that raise the same kind of concern over the quality and content of online conversations and public debate. All these compounded risks threaten to distort political debate, affect democracy, and even endanger public safety.[9] Additionally, OpenAI reported an estimated 100 million active users of ChatGPT in January 2023, which makes the potential for a vast and systemic impact of these risks a considerable one.[10]

LLMs are also expected to have great potential. They will transform a variety of industries, freeing up professionals' time to focus on different substantive matters. They may also improve access to various services by facilitating the production of personalized content, for example for medical patients or students.[11] Consequently, one of the key policy questions LLMs pose is how to regulate them so that some of these risks are mitigated while still encouraging innovation and allowing their benefits to be realized.[12] This Essay examines this question, with a focus on the liability regime for LLMs for speech and informational harms and risks in the European Union.

The EU is undertaking an ambitious regulatory project to pursue a digital transformation "that works for the benefit of people through respecting our values."[13] A central part of this effort is the proposed Artificial Intelligence Act (AI Act), a flagship risk-based regulation of trustworthy AI. The AI Act would be a Europe-wide

---

[8] *See* Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI*, 41 COMP. L. & SECURITY REV. 105567 (2021).

[9] *See also* Melissa Heikkila, *The Algorithm*, MIT TECH. REV. (Apr. 3, 2023), which discusses the cybersecurity risks.

[10] Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base—Analyst Note*, REUTERS (Feb. 2, 2023), https://perma.cc/4Q6L-HPJT].

[11] *See* DonHee Lee & Seong No Yoon, *Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges*, 18 INT'L J. ENVTL. RES. & PUB. HEALTH 271 (2021); Alberto Nasciuti, *Exploring the MadTech & AI Frontier: Unveiling the LLM Revolution!* LINKEDIN (June 8, 2023), https://perma.cc/3RF7-R9ZA].

[12] *See* Chloe Xiang, *The Open Letter to Stop 'Dangerous' AI Race Is a Huge Mess*, VICE (Mar. 29, 2023), https://perma.cc/6QCC-F3YN.

[13] EUROPEAN COMMISSION, SHAPING EUROPE'S DIGITAL FUTURE, COM (2020) 67 final (Feb. 19, 2020), https://perma.cc/R9JB-LBA6.

law designed to address some of the "traditional" and long-identified ethical risks posed by AI systems, such as lack of technical and legal transparency, the potential for bias and discrimination, and danger to privacy.[14] The Act splits AI systems into four different levels of risk, prohibits a limited set of systems that pose an unacceptable level of these risks (such as real-time remote biometric identification systems in publicly accessible spaces), and is mostly concerned with creating obligation for the second tier of risk, "high risk" systems.[15] High risk systems are a limited set of systems that acutely raise these kinds of risks, judging by their intended use as determined by their designer, but to a degree that can be mitigated. The Act then creates a variety of safety requirements for such systems related to data governance, transparency, and their design and operation. It requires, for example that high-risk systems be supervised by a human when in use.[16] As it turns out, however, the AI Act, which is still being discussed at the time of writing, appears to be rather ill-prepared to address some of the challenges raised by LLMs, like ChatGPT.[17] It is not intended to address the risks of systems when it is the user who determines how the systems are to be used, as is the case with general purpose AI systems. It is also not intended to address content moderation, freedom of expression, or information-related harms and risks.[18]

---

[14] *See, e.g.,* Brent Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter & Luciano Floridi, *The Ethics of Algorithms: Mapping the Debate*, 3(2) BIG DATA & SOCIETY (2016).

[15] *See* Lilian Edwards, *The EU AI Act: A Summary of Its Significance and Scope*, ADA LOVELACE INSTITUTE 9 (Apr. 8, 2022), https://perma.cc/NNR2-T4SK.

[16] Natali Helberger & Nicholas Diakopoulos, *ChatGPT and the AI Act*, 12 INTERNET POL'Y REV. 1, 3 (2023).

[17] *See* Philipp Hacker, Andreas Engel & Marco Mauer, *Regulating ChatGPT and other Large Generative AI Models* (working paper, revised 12 May 2023), https://perma.cc/AK8E-3SD9 [hereinafter Hacker, Engel & Mauer, *Regulating Chat GPT May 2023 version*].

[18] At the time of writing, different amendments are being discussed that seek, in different ways, to extend the obligations of high-risk models to LLMs. Scholars like Philip Hacker et al. have argued that these approaches are still unconvincing because they focus on foundation models alone, rather than the use case or specific applications. Extending the obligations of high risk models to all LLMs may also prove unworkable: "Providers of LGAIMs such as ChatGPT would, therefore have to analyze the risks for every single, possible application in every single high-risk case . . . . This seems not only almost prohibitively costly but also hardly feasible. The entire analysis would have to be based on an abstract, hypothetical investigation, and coupled with—again hypothetical—risk mitigation measures that will, in many cases, depend on the concrete deployment, which by definition has not

The European Union, however, recently enacted another regulation that is directly concerned with addressing the risks and challenges associated with content moderation, freedom of expression and the spread of disinformation or other forms of harmful speech online: the Digital Services Act (DSA).[19] The challenge, however, is that the DSA was not meant to cover AI generated content, but rather user generated content.[20] At first sight it thus does not seem to apply to the content generated by LLMs.[21]

This Essay argues, however, that because many of the risks these systems raise are risks to the information ecosystem, in Europe they can and should be addressed, at the outset, with current content moderation law. This Essay proposes an interpretation of the DSA that could apply to these tools when they are released in the market in a way that strongly resembles other intermediaries covered by content moderation laws, such as search engines. (This is without prejudging present and future AI regulations that may be created to deal with other challenges in a more specific way.) In doing so, it follows other scholars who have argued that the regulation of LLMs should focus on concrete risks they entail based on their specific uses.[22] In the US and elsewhere, it may be helpful to use an approach that mixes the traditional safe harbor for internet intermediaries with due-diligence and risk-mitigation obligations, especially for the largest providers. Such an approach has the potential to, on the one hand, continue to support innovation, research, and development, while on the other hand, create incentives for innovation to be done responsibly, and in a way that mitigates potential systemic risks and harms.

This Essay's point of departure is the DSA, Europe's main content moderation law. The DSA is a functional analog to 47 U.S.C. § 230 and it updated the generally

---

been implemented at the moment of analysis." *See* Hacker, Engel & Mauer, *Regulating Chat GPT May 2023 version, supra* note 17, at 5; Luca Bertuzzi, *AI Act Enters Final Phase of EU Legislative Process,* EURACTIV (June 14, 2023), https://perma.cc/9Y2J-EQJ3.

[19] Regulation of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act). PE (2020) 825 final (Dec. 15, 2020), Article 1 [hereinafter DSA].

[20] Luca Bertuzzi, *Leading EU Lawmakers Propose Obligations for General Purpose AI*, EURACTIV (Mar. 14, 2023), https://perma.cc/A769-X747].

[21] Hacker, Engel & Mauer, *Regulating Chat GPT May 2023 version*, *supra* note 17, at 17; Helberger & Diakopoulus, *supra* note 16.

[22] *See, e.g.,* Hacker, Engel & Mauer, *Regulating Chat GPT May 2023 version, supra* note 17.

applicable ground rules for the regulation of online content. It seeks to balance different European values like protecting freedom of speech and information, maintaining high levels of consumer protection and fostering innovation and economic growth.[23] As I explain in Part II, it does this by providing a somewhat conditional safe harbor from liability for all platforms hosting content, while also imposing on the most systemically relevant actors—like very large social media companies and search engines—certain due diligence and risk-mitigation obligations to attenuate some of the systemic threats that these tools pose.[24] An advantage of this approach is that it simplifies many of the difficulties of imposing liability for AI-related harms. And the DSA is already law; it will take at least a few years until the AI Act enters into full force.[25]

Contrary to what may happen in the United States, where there is at least disagreement on whether tools like ChatGPT are covered by Section 230 or not,[26] scholars in the EU do not think that LLM-chatbots fall naturally under the EU safe harbor for intermediary liability.[27] The main reason is that large language models generate content themselves, and the definition of exempt intermediaries that could best fit ChatGPT, a hosting service, refers to "content provided by users."[28]

If this is adopted as the main interpretation, companies deploying interfaces that provide end-users easy access to LLMs that may generate harmful and illegal disinforming or defamatory content could be held liable for such content.[29] Far from being a panacea, such liability could undermine research, development, and innovation by creating legal uncertainty for businesses developing and adopting various generative AI systems. No one really knows how to train these powerful AI

---

[23] *See* DSA, *supra* note 19, Recitals 1, 3.

[24] *See infra* Part II.

[25] Lilian Edwards, *Can the EU AI Act Successfully Regulate Generative AI*, Presentation at Sciences Po Law School (Mar. 30, 2023).

[26] *See* Derek E. Bambauer & Mihai Surdeanu, *Authorbots*, 3 J. FREE SPEECH L. 375 (2023); Matt Perault, *Section 230 Won't Protect ChatGPT*, 3 J. FREE SPEECH L. 363 (2023).

[27] Philipp Hacker, Andreas Engel & Theresa List, *Understanding and Regulating ChatGPT, and Other Large Generative AI Models: With Input from ChatGPT*, VERFBLOG (Jan. 20, 2023), https://perma.cc/EWD7-XZRV.

[28] *Id.*

[29] *See, e.g.,* Reuters, *Australian Mayor Prepares World's First Defamation Lawsuit over ChatGPT Content*, GUARDIAN (Apr. 6, 2023), https://perma.cc/3MLU-QYSW.

systems so that they will always be reliable, helpful, honest, and harmless.[30] At the same time, there are many actors involved in the chain of events and the training of an algorithm (designers, manufacturers, deployers, users), AI systems are opaque, and AI systems can fail (and harm people) in unpredictable ways. How to allocate fault amongst the different actors is often unclear, and proving that someone breached a duty of care can be very hard for victims.[31] To avoid some of these concerns, regulators both in the EU and the US are adopting different risk regulation mechanisms—for creating *ex ante* requirements like conducting risk assessments and following technical standards—to deal with AI and social media. I discuss this further in Section III. [32]

In this Essay I start proposing a middle-ground position in which general purpose LLMs like ChatGPT, Bard, and LLaMA should be, and perhaps already are, covered by internet intermediary regulation. At the same time, the companies placing these systems on the market should also be required to comply with due diligence and risk-mitigation obligations to, for example, take measures to curb harmful speech. This would achieve a balance between facilitating the development of new tools and services while ensuring that their creators set in place key guardrails before placing them on the market. This is the DSA approach to social media regulation in Europe. And lawyers and policymakers in the United States may also want to consider this, both because these EU regulations are applicable to US companies operating within the EU, and because the US is also moving in the direction

---

[30] *See* Urs Gasser, *Zur Forderung Eines Moratoriums für die KI-Entwicklung 'Eine Pause Beim Training von Künstlicher Intelligenz Hilft Nicht'*, TUM (Apr. 3, 2023), https://perma.cc/KJ4Z-PR44.

[31] *See* Miriam Buiten, Alexandre de Streel & Martin Peitz, *The Law and Economics of AI Liability*, 48 COMPUTER L. & SECURITY REV. 150794, 6 (2023); Christiane Wendehorst, *AI Liability in Europe: Anticipating the EU AI Liability Directive*, ADA LOVELACE INSTITUTE (Sept. 22, 2022); *see also* OpenAI, *How Should AI Systems Behave, Who Should Decide?,* OPENAI, https://perma.cc/6A3T-U4UR; Margot Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. __ (forthcoming 2023).

[32] Kaminski, *supra* note 31; *see also* Wendehorst, *supra* note 31. It is also worth noting that the EU Parliament is discussing an AI Liability Directive that may ease some of these challenges. Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive) 2022/0303(COD), COM(2022) 496 final (Sept. 28, 2022).

of AI risk regulation and the future of Section 230 is still part of the political agenda.[33]

Specific to the EU context, I offer an alternative interpretation and legal reform proposal for the DSA, hoping that EU courts and scholars will not dismiss, just yet, the question of whether the DSA applies to ChatGPT or other LLMs. I propose a functional and teleological interpretation of the DSA, one in which courts, lawyers, and lawmakers should consider the intention behind the DSA, the way in which new intermediaries are being used, and the function they serve in the information environment. It is an open secret that even if OpenAI and Google continue to label these bots as experiments, most people—myself included—have spent the last few months using ChatGPT to replace or complement their search engines.[34] This is not unreasonable: the landing page of ChatGPT, for example, invites users to ask questions on topics ranging from quantum computing to children's birthday party ideas—normal queries one would use a search engine for.[35] Thus, I suggest that if these tools are being placed on the market where they can be functionally and reasonably considered to be used for search purposes, they should then be bound to the same safe harbor and due-diligence risk mitigation obligations as search engines and other online platforms.

To lay out this argument in more detail, this Essay proceeds as follows: Section I provides background on the EU's content moderation framework, focusing on the DSA and illegal speech liability. Section II explains why, from a policy perspective, it is difficult, and perhaps undesirable, to solely rely on member states' online speech laws and intermediary liability rules when dealing with LLM-generated harms, and why a risk-regulation approach could be desirable. Section III explains how the DSA could be interpreted to apply to LLM-powered general information retrieval tools.

---

[33] *See, e.g.,* Brian Fung, *Senators Warn Big Tech on Section 230: 'Reform Is Coming'*, CNN BUSINESS (Mar. 8, 2023), https://perma.cc/VA6X-HD3R; Tate Ryan-Mosley, *Four Ways the Supreme Court Could Reshape the Web*, MIT TECH. REV. (Feb. 27, 2023), https://perma.cc/ALN9-XF2P. Additionally, the EU has become a standard setting institution in regard to regulating technology, a phenomenon that is known as the Brussels Effect. *See* ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD (2020).

[34] David Pierce, *Google Says Its Bard Chatbot Isn't a Search Engine—So What Is It?*, THE VERGE (Mar. 21, 2023), https://perma.cc/8YF4-BH3T.

[35] ChatGPT, https://chat.openai.com/ (last visited June 15, 2023).

I. CONTENT MODERATION LAW TO PRODUCE A HEALTHIER INFORMATION ENVIRONMENT: THE EU APPROACH

This first section briefly presents the EU content moderation framework, with a focus on how European policymakers think about freedom of speech and information, and then presents the DSA's intermediary liability framework.

A. The Right to Freedom of Expression and Information Under EU Law

The European Regulatory Framework on Freedom of Expression is centered around Article 10 of the European Convention of Human Rights, a treaty that binds all EU Member States, and Article 11 of the European Charter of Fundamental Rights, the European Union's flagship instrument for protecting human rights. Freedom of expression and information is recognized as a general principle of EU law and as a fundamental right. Consequently, all EU and member state legislation must fully respect and comply with these principles.[36] These Articles, the latter one modeled after the first one,[37] uphold freedom of expression, which includes the right to receive and impart information, and a commitment to respect media freedom and pluralism as pillars of modern democracy and enablers of free and open debate.[38]

In the EU, freedom of expression and information are strongly protected. However, it is also understood that these freedoms carry with them duties and responsibilities for both public and private actors, which are expressed as conditions, restrictions, and penalties prescribed by law.[39] The European Court of Human Rights (ECtHR) has developed extensive case law on the scope of these rights and

---

[36] Treaty on European Union Article 6.1, Nov. 1, 1993, O.J. (C 224) 1 (consolidated version 2016).

[37] Article 11 of the Charter is modelled after Article 10 of the Convention, and the Charter itself provides that inasmuch as the Charter contains rights that correspond to those in the ECHR, "the meaning and scope of those rights shall be the same as those laid down by the Convention." Tarlach McGonagle, *Free Expression and Internet Intermediaries: The Changing Geometry of European Regulation*, in OXFORD HANDBOOK OF ONLINE INTERMEDIARY LIABILITY (Giancarlo F. Frosio ed., 2020).

[38] *See* European Commission, *Media Freedom and Pluralism*, https://perma.cc/MR4D-WWQX.

[39] *See, e.g.,* Thomas Hochmann, *Why Freedom of Expression is Better Protected in Europe than in the United States*, 2 J. FREE SPEECH L. 63 (2023).

limitations.[40] In general, it has only permitted restrictions where strictly necessary to sustain other public interest goals like democracy, national security, territorial integrity, public safety, or protection of the reputation rights of others.[41] Restrictions must satisfy the ECtHR's standard four-step test: being legally prescribed, pursuing a legitimate aim, being necessary in a democratic society, and being a proportionate means to reach the aim pursued.[42] The European Court of Justice applies a slightly different test: limitations must be in the public interest, be laid out in law, and be proportionate.[43] Member states regulate online speech through their national civil and criminal laws on hate speech and other forms of illegal speech, within these general parameters.[44]

With the advent of the Internet, the ECtHR has identified a positive obligation for states to guarantee media pluralism, and to endeavor to create a positive environment for participation in public debate without fear. In *Dink v. Turkey*, the court ruled that Turkish authorities had failed in their positive obligation to protect the life of Hrant Dink, an Armenian journalist, despite knowing that he had received serious death threats. The court concluded that this failure had a detrimental effect on the exercise of freedom of expression in Turkey.[45] Commentators have argued that this positive dimension could also involve requiring states to adopt measures

---

[40] The European Court of Human Rights is the adjudicatory body formally tasked with interpreting the Convention, which binds all forty-seven Member State signatories of the EU Charter.

[41] Article 10 of the European Convention on Human Rights (ECHR). The European Charter of Fundamental Rights is a document that sets out the fundamental rights within the European Union. The European Convention on Human Rights is an international treaty that protects human rights across Europe and is incorporated into the legal systems of the EU member states. Pursuant to Article 52(3) of the Charter, the meaning and scope of the right to freedom of expression are the same as those guaranteed by the ECHR.

[42] McGonagle, *supra* note 37.

[43] Charter of Fundamental Rights of the European Union Article 51(3), [O.J.] C 303/1 (2010).

[44] *See* Beatriz Botero Arcila & Rachel Griffin, *Social Media, Fundamental Rights, Democracy and The Rule of Law* (European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs 2023) (showing how some rules within the EU on disinformation are still problematic from a fundamental rights perspective).

[45] Dink v. Turkey, Judgment of Sept. 14, 2010, applications nos. 2668/07, 6102/08, 30079/08, 7072/09 and 7124/09, Eur. Court H.R. (ser. A) No. 219, 2010-VIII.

to foster media pluralism or adopt measures to regulate the moderation of disinformation online.[46]

B. Very Basic Internet Intermediary Liability in EU Law and Content Moderation Law

The baseline regime for internet intermediary governance in the EU was established by the 2000 E-Commerce Directive. In 2022, the DSA replaced the E-Commerce Directive, and the DSA is now the most important legal instrument setting out generally applicable ground rules for the regulation of online content. The idea of replacing the E-Commerce Directive was to address some of the new risks and challenges raised by the different business models that have emerged and spread since the Directive was enacted. These encompass social media, online platforms, search engines, and cloud infrastructure services.[47]

The DSA maintained the baseline principle of content regulation that had been established by the E-Commerce Directive: internet intermediaries are exempt from liability for hosting content that is illegal, so long as they (1) do not participate in its production, and (2) remove it once made aware of it. The DSA also created incentives for intermediaries to engage in co- and self-governance approaches and created a few new substantive obligations for online intermediaries. Importantly, very large online platforms and very large online search engines must now conduct periodic risk assessments and must design risk mitigation measures accordingly, given their reach and their central role in the information ecosystem.[48] Very large online platforms and search engines are classified as those that have over 45 million users in the EU.[49] This approach seeks to balance different European values, including protecting and guaranteeing fundamental rights and freedoms such as the freedom of speech and information, maintaining high levels of consumer protection, and maintaining and strengthening economic growth and innovation.[50]

---

[46] Judith Bayer et al., *The Fight Against Disinformation and the Right to Freedom of Expression*, PE 695.445 3, at 21 (European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs, 2021).

[47] DSA, *supra* note 19, Recitals 1, 28.

[48] *Id.* Recital 59.

[49] *See* European Commission, *DSA: Very Large Online Platforms and Search Engines* (last updated Apr. 25, 2023), https://perma.cc/47CZ-7A25.

[50] DSA, *supra* note 19, Recitals 1, 3.

1. The conditional safe harbor for online intermediaries

The key principle of internet intermediary governance in the EU is that intermediary services are exempt from liability for hosting, distributing, or transmitting illegal content so long as they do not participate in its production and they remove illegal content as soon as they are made aware of it.[51] Intermediaries include services as varied as social media,[52] internet service providers,[53] search engine advertising services,[54] and online sales platforms.[55] As in the United States, the central ideas behind this exemption from liability are (1) to help innovation and sustain the growth of the internet and the digital economy,[56] and (2) to avoid undue platform interference with privacy and with freedom of expression and information, since otherwise platforms would have strong incentives to closely surveil and censor users to prevent any potentially illegal activity.[57] In the EU, intermediary liability was also the basis for harmonizing national laws on the legal responsibility of internet service providers and aiding in the development of the internal digital market.[58] The DSA maintained this but updated and expanded it with new obligations which aim to address the dissemination of illegal content on platforms while protecting users' fundamental rights. The DSA was implemented on November 2022 and will be fully applicable across the EU in early 2024.[59]

---

[51] *Id.* Recitals 16–18.

[52] Netlog, C-360/10, 2012 E.C.R. I-0000 (ECJ Mar. 16, 2010).

[53] *Id.*

[54] Google France, C-236/08, 2010 E.C.R. I-2417 (ECJ Mar. 23, 2010).

[55] L'Oréal, C-324/09, 2011 E.C.R. I-0000 (ECJ June 12, 2011).

[56] *See* ANDREJ SAVIN, EU INTERNET LAW, THIRD EDITION., ELGAR EUROPEAN LAW (2020).

[57] Folkert Wilman, *The EU's System of Knowledge-Based Liability for Hosting Service Providers in Respect of Illegal User Content—Between the e-Commerce Directive and the DSA,* 12 JIPITEC 317 (2021).

[58] Joris van Hoboken, Joao P. Quintais, Joost Poort & N. van Eijk, *Hosting Intermediary Services and Illegal Content Online: An Analysis of the Scope of Article 14 ECD in Light of Developments in the Online Service Landscape* 28 (European Comm'n 2018).

[59] Operators designated as very large online platforms (VLOPs) and very large online search engines (VLOSEs) will have to comply with stricter obligations already from mid-2023, *see* European Parliament, *Digital Services Act: Application Timeline* (Think Tank, European Parliament, Nov. 30, 2022), https://perma.cc/3KDA-QBN2.

The DSA applies to online intermediary services, which it broadly defines as "information society services"[60] that are either a "mere conduit service," a "caching" service, or a "hosting" service.[61] Information society services are defined as "any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient,"[62] essentially meaning a commercial service predominantly provided via the internet. Mere conduit services and caching services "of a mere technical, automatic, and passive nature,"[63] which merely transmit information for users or temporarily store information to make transmission more efficient.[64] The protection is thus granted when a service is (1) an information society service and, if so, (2) when the activity falls into one of those three categories.[65]

Most platforms are hosting services. Hosting services and their exemption from liability are defined in Article 6 of the DSA as follows:

> Article 6 Hosting
>
> 1. Where an information society service is provided that consists of the storage of information provided by a recipient of the service, the service provider shall not be liable for the information stored at the request of a recipient of the service, on condition that the provider:
>
> (a) does not have actual knowledge of illegal activity or illegal content and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or illegal content is apparent; or
>
> (b) upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the illegal content.

---

[60] Information society services is a term used in EU law to refer, very broadly, to companies providing internet services and services through the Internet. Information society services are defined in Directive 2015/1535 which lays out a variety of standards on the provision of information services, as "any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services." Directive (EU) 2015/1535, Article 1(b).

[61] DSA, *supra* note 19, Article 3(g).

[62] Directive (EU) 2015/1535 of the European Parliament and of the Council OJ L 241, 17.9.2015, art. 1.1(b); *see also* DSA, *supra* note 19, Recital (5).

[63] *See* Tobias Mc Fadden v. Sony Music Entertainment Germany GmbH, Case C-484/14, [2016] E.C.R. I-0000 (ECJ Sept. 15, 2016).

[64] DSA, *supra* note 19, Article 5.

[65] Jennifer Cobbe & Jatinder Singh, *Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges*, 42 COMPUTER L. & SECURITY REV. 105573, 39 (2021).

The basic definition of hosting services, in paragraph 1, thus applies to many intermediary functions in today's online environments. It includes web hosting, online media sharing platforms (like YouTube or Bandcamp), file storage and sharing platforms (like Dropbox), social media, and video game platforms (such as Xbox Live and World of Warcraft).[66] The immunity from liability for hosted content is conditioned upon the service providers not having actual knowledge of illegal activity or content[67] and on their removing illegal material promptly once they have been made aware of its presence on the platform.[68] If the DSA is interpreted as the E-Commerce Directive was—which is not unlikely as its wording is the same—service providers who do not meet these conditions may be liable for hosting illegal content.[69]

To avoid state censorship and surveillance, the DSA forbids Member States from imposing a general obligation to monitor for illegal content or activity.[70] As to freedom of expression and its limitations, the DSA defers to national law to define illegal speech.[71]

---

[66] Van Hoboken et al., *supra* note 58, at 13, 14.

[67] *See also id.* at 29 (brackets in original), clarifying that in the E-commerce Directive the equivalent provision contained two standards related to the illegal activity or information stored, "[potentially referring to two types of wrongdoings]: (i) 'actual knowledge' and (ii) 'awareness of facts or circumstances' from which the illegality is 'apparent,' also referred to as 'constructive' or 'construed' knowledge. The *travaux préparatoire*s of the ECD appear to support this distinction, with the result that criminal liability of hosting platforms would require actual knowledge on the part of the hosting service provider, whereas civil liability regarding claims for damages would require solely constructive knowledge."

[68] As explained in more detail below, the provision adds conditions that the hosting service provider must meet to be exempt from liability: the provider (a) must "not have actual knowledge of illegal activity or illegal content and, as regards claims for damages" and not be "aware of facts or circumstances from which the illegal activity or illegal content is apparent," and (b) "upon obtaining such knowledge or awareness, must act[] expeditiously to remove or to disable access to the illegal content." DSA, *supra* note 19, Article 8.

[69] DSA, *supra* note 19, Article 3(t).

[70] *Id.*

[71] *See also* Botero Arcila & Griffin, *supra* note 44 (discussing how this has several benefits but also raises concerns related to the adequacy of fundamental rights protection).

2. Self-regulation and co-regulation approaches

Illegal speech is not the only content that platforms must or are encouraged to remove in Europe. Platforms can also set and enforce standards regarding what content they will permit. Content moderation refers broadly to the activities and measures platforms take to detect and address illegal content and information incompatible with their terms of service.[72] Platforms have been under regulatory, economic, and reputational pressures from governments and other stakeholders because of the direct effects their content moderation practices and policies have on enabling or restricting freedom of expression.[73] OpenAI also has its own content policy and has set in place a sophisticated content "moderation endpoint" able to detect undesired content. As described by OpenAI,

> When given a text input, the Moderation endpoint assesses whether the content is sexual, hateful, violent, or promotes self-harm—content prohibited by our content policy. The endpoint has been trained to be quick, accurate, and to perform robustly across a range of applications. Importantly, this reduces the chances of products "saying" the wrong thing, even when deployed to users at-scale.[74]

The DSA also encourages platforms to draw up voluntary common codes of conduct and agreements to tackle different types of illegal content and systemic risks, such as the Code of Practice on Disinformation or a Code of Conduct on Countering Illegal Hate Speech Online. These are documents in which platforms commit to conduct such as taking more action on manipulative behavior or reviewing reported hate speech in under 24 hours.[75] For systemic risks, as I explain in detail below, the DSA makes some of these agreements and their goals binding.[76]

---

[72] *Id; see, e.g.,* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018).

[73] Robert Gorwa, *Stakeholders*, PLATFORM GOVERNANCE TERMINOLOGIES ESSAY SERIES-YALE-WIKIMEDIA INITIATIVE ON INTERMEDIARIES & INFORMATION (2022).

[74] *New and Improved Content Moderation Tooling,* OPENAI (Aug. 10, 2022), https://perma.cc/FU9P-V2Z7.

[75] European Commission, *Code of Practice on Disinformation* (Dec. 5, 2018), https://perma.cc/X76N-QBPR; European Commission, *Code of Conduct on Countering Illegal Hate Speech Online* (May 31, 2016), https://perma.cc/KR4J-7XLD.

[76] DSA, *supra* note 19, Article 45.

3. Due diligence and risk mitigation obligations

The key innovation brought about by the DSA is a set of due diligence obligations aimed at ensuring a safe and transparent online environment.[77] These obligations are tailored to the different types and sizes of the intermediary services.[78] All providers, for example, now have a duty to publish a yearly report "on any content moderation that they engaged in during the relevant period."[79] All hosting services must enable a mechanism through which people can notify them "of the presence on their service of specific items of information that the individual or entity considers to be illegal content."[80] Hosting services must also institute an accessible complaint-handling system through which users whose content was removed, or who have reported content which was not removed, can appeal the decision.[81]

The providers of very large hosting services are considered to pose a societal systemic risk "stemming from the design, functioning and use of their services, as well as from potential misuses by the recipients of the service," and are subject to stricter obligations. The DSA refers to these kind of services as very large online platforms and very large online search engines.[82] Systemic risks, as identified by the Act, include (1) risks associated with the dissemination of illegal content,[83] (2) risks of negative effects for the exercise of fundamental rights, such as human dignity, freedom of expression and information, data protection, and the right to nondiscrimination,[84] (3) and risks concerned with the actual or foreseeable effects of the platforms on democratic processes and civic discourse.[85] Platforms must also consider how these risks can be influenced by intentional manipulation of platforms'

---

[77] *Id.* Recital 40.

[78] *Id.* Recitals 40, 41.

[79] *Id.* Article 15.

[80] *Id.* Article 16 (establishing that users' reporting potentially illegal content is considered to give rise to actual knowledge or awareness about the presence of such content).

[81] *Id.* Article 20.

[82] *Id.* Recital 79; *see also* European Commission, *supra* note 45.

[83] DSA, *supra* note 19, Recitals 80, Article 34.1(a).

[84] *Id.* Recital 81 (noting that "[s]uch risks may arise, for example, in relation to the design of the algorithmic systems used by the very large online platform or by the very large online search engine or the misuse of their service through the submission of abusive notices or other methods for silencing speech or hampering competition"), Article 34.

[85] *Id.* Recital 82.

services (for example, through inauthentic use or automated exploitation of the services, or the amplification and rapid dissemination of illegal content).[86] According to the Act, very large online platforms and very large search engines will now have to conduct regular risk assessments on the systemic risks detailed above,[87] and take reasonable, proportionate, and effective measures to mitigate the risks identified in their impact assessments. The risk assessments and measures must be independently audited every year and will generally be overseen by the Commission.[88]

II. THE DIFFICULTY OF RELYING ON MEMBER STATES' ONLINE SPEECH LAWS AND LIABILITY RULES WHEN DEALING WITH LLMS: TOWARDS RISK MITIGATION

A technical report published by OpenAI in late March 2023 outlines several of the safety risks that arise from ChatGPT-4 and explains what OpenAI did to mitigate those risks. The self-reported risks include generating harmful content, hallucinations,[89] disinformation that can be used in influence operations, material that undermines privacy and cybersecurity,[90] and material that "reinforce[s] and reproduce[s] specific biases and worldviews, including harmful stereotypical and demeaning associations for certain marginalized groups."[91] The report also provides examples from prompts that OpenAI's technical team used in testing before training it to refuse some of these requests. The prompts included asking how to kill the most people with only $1, how to avoid detection while laundering money, how to write a "very explicit letter threatening someone with gang rape" and how to write "I hate Jews but in a way that would not get taken down by Twitter."[92] Before release, the system answered all these questions in detail. OpenAI's team then trained

---

[86] *Id.* Article 34.2.

[87] *Id.*

[88] *Id.* Article 37.

[89] According to the report, "GPT-4 has the tendency to 'hallucinate,' i.e. 'produce content that is nonsensical or untruthful in relation to certain sources.' This tendency can be particularly harmful as models become increasingly convincing and believable, leading to overreliance on them by users. . . . Additionally, as these models are integrated into society and used to help automate various systems, this tendency to hallucinate is one of the factors that can lead to the degradation of overall information quality and further reduce veracity of and trust in freely available information." OpenAI, *Technical Report, supra* note 2, at 46.

[90] *Id.* at 44.

[91] *Id.* at 47.

[92] *Id.*

the system for refusal in these and other instances, which was effective in most cases.[93]

The report highlights, however, that training for refusal also has certain limits: as to the antisemitic question, the system still generated an answer (even if a tame one).[94] In some other instances, refusals and other mitigations measures can exacerbate bias by yielding a higher number of false positives. This can occur due to bias in the classifiers, for example when content that pertains to minorities—such as LGBTQ couples—is labelled as "harmful" or "toxic."[95] Lastly, a key limitation is that users have devised workarounds to content moderation safeguards.[96]

Most of the results that were generated before release could be considered illegal speech in a variety of European countries, as are some of the results reportedly generated by ChatGPT when users bypass content moderation.[97] It is not totally clear yet, however, who should be liable for creating such content, as in general, it is difficult to assign liability for AI-generated harms. This happens for several reasons:

First, AI systems are opaque, which makes it hard to determine causality: It is often unclear how input resulted in output. Some harms—for example if an LLM system gives the wrong medical advice to a patient—may not be obvious for injured parties to identify, or to trace back the source of the harm.[98] Additionally, understanding AI systems requires technological expertise, raising the costs of litigation.[99]

---

[93] *Id*. at 48.

[94] *Id.*

[95] *See id.* at 49 (citing Albert Xu, Eshaan Pathak, Eric Wallace, Suchin Gururangan, Maarten Sap & Dan Klein, *Detoxifying Language Models Risks Marginalizing Minority Voices*, in PROCEEDINGS OF THE 2021 CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES 2390–97 (2021)).

[96] *See, e.g.,* Josh Taylor, *ChatGPT's Alter Ego, Dan: Users Jailbreak AI Program to Get Around Ethical Safeguards*, GUARDIAN (Mar. 8, 2023), https://perma.cc/4UQX-VHYX.

[97] *Id.; see also* GIOVANNI PITRUZZELLA & ORESETE POLLICINO, DISINFORMATION AND HATE SPEECH: A EUROPEAN CONSTITUTIONAL PERSPECTIVE (2020).

[98] Buiten et al., *supra* note 31, at 99.

[99] Kaminski, *supra* note 31, at 18; Wendehorst, *supra* note 31.

Second, AI is complex because many stakeholders and components are involved in its development and deployment. Thus, who is the person who operates or controls the system is by no means an easy question: Many actors—designers, manufacturers, deployers, users—are involved in the chain of events and the training of an algorithm that can lead to a potential instance of harm; how to allocate fault among them is not very clear. Jennifer Cobbe and Jatinder Singh have shown this in the context of AI as a service (AIaaS), where complex networks and dynamic relationships between AIaaS providers and customers don't quite fit traditional data protection roles as data controller and processor.[100] OpenAI foresees, for example, that it will "pre-train" and "fine-tune" these models, but that these will then be "customizable by each user up to limits defined by society."[101] It will be hard to determine to what extent providers, customers, and even end-users will be considered liable for content-related harms in this scenario.

Christiane Wendehorst has argued that the difficulty in identifying who controls a system may lead to an accountability gap because, in most countries, vicarious liability rules do not apply to machines. Thus, when tasks that were traditionally carried out by a human are now carried out with or by an AI and they lead to a harm, it is possible that no one is liable at all.[102] The AI system itself cannot be liable; the deployer is not liable if they demonstrate that they bought or acquired the AI system from a recognized provider, and can prove that they complied with monitoring duties; and the producer is often not liable because proving a defect in an AI system is hard.[103] Relatedly, the European Union's Expert Group on Liability and New Technologies has pointed out that the liability laws of Member States, as they stand today, tend to assume the human who is in the front end of the system—such as the security driver of an automated vehicle—is in full control of the vehicle and may thus be responsible for an accident. [104] This, however, is problematic because

---

[100] Cobbe & Singh, *supra* note 65, at 8.

[101] OpenAI, *supra* note 31.

[102] *See* Kaminski, *supra* note 31; Wendehorst, *supra* note 31.

[103] Christiane Wendehorst, *Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks*, in THE CAMBRIDGE HANDBOOK OF RESPONSIBLE ARTIFICIAL INTELLIGENCE: INTERDISCIPLINARY PERSPECTIVES 187–209 (Silja Voeneky et al. eds., 2022).

[104] EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES—NEW TECHNOLOGIES FORMATION, LIABILITY FOR ARTIFICIAL INTELLIGENCE AND OTHER EMERGING DIGITAL TECHNOLOGIES (2019),

it tends to leave technology companies off the hook, while the research shows that humans tend to trust AI systems.[105]

Take the following example: In 2016, Twitter users taught a Microsoft AI chatbot to be racist in less than a day.[106] In situations where the racist statement falls within the scope of a European ban on discrimination or hate speech, would the users be liable? They may be—and most likely they should be if they were interacting with the system maliciously. But if liability would fall *only* on the users who interacted maliciously, then Microsoft would have no due diligence obligation to try to avert those attacks.

Third, though lawmakers could impose much stricter liability on developers, such liability may be unsuited to the special features of AI systems and may have limited effects: One of the characteristic features of AI systems is their complexity, through which they can produce outcomes that they were not explicitly programmed to produce. Consequently, not all AI harms will be foreseeable by human programmers from the outset, which make such harms harder to disincentivize through litigation. For this reason, and assuming that AI systems may have an important value for society, authors like Miriam Buiten et al. have argued that a strict liability standard may be an excessive burden.[107]

Lastly, and particular to the EU, a solely liability-based regime to address many of these harms would leave litigants in a regulatory fragmented environment. Different member states not only define illegal speech differently, but also have different tort law doctrines that apply to actors whose omissions or actions contribute to harm.[108] This may be hard to navigate for plaintiffs, but it also raises concerns about legal uncertainty for businesses that could hamper innovation within the EU.[109]

---

https://perma.cc/VU3R-C5A8; *see also* Ryan Calo, *Robots in American Law* (U. Wash. Sch. L. Legal Stud. Rsch. Paper No. 04, 2016), at 24 showing how, in the United States, courts tend to assign fault to the human right at the end.

[105] *See* Ben Green & Amba Kak, *The False Comfort of Human Oversight as an Antidote to A.I. Harm*, SLATE (June 15, 2021), https://perma.cc/9QLY-H4H4.

[106] *See* James Vincent, *Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less Than a Day*, THE VERGE (March 24, 2016), https://perma.cc/G9GF-SUVX.

[107] Buiten et al., *supra* note 31.

[108] *See also* Wendehorst, *supra* note 31 (explaining this and clarifying how it may, to some limited extent, be addressed by the Proposed AI Liability Directive).

[109] *Id.*

These and other regulatory challenges have led scholars and policymakers to argue that ex-post methods of regulation like liability are not enough to manage the risks associated with AI systems.[110] AI regulation has started adopting tools of risk regulation in both Europe and in the United States.[111] In the EU, this began with the General Data Protection Regulation (GDPR), which is now also an important element of the AI Act and, as we saw in the previous Part, of the DSA. The advantage of this approach is that it incentivizes the providers and developers of these tools to think about the safety of these systems from the beginning. There are downsides to this approach too: The approach includes a tacit acceptance that some risks will indeed materialize, a preference for previously known and quantifiable harms, and a de-emphasizing of making injured people whole as a priority.[112] Some of these regulatory tools could be complemented with adequate mechanisms of individual redress, though the details of that go beyond the scope of this article.[113]

From a policy perspective it would make sense for companies like OpenAI, DeepMind, Nvidia, Meta, and others placing LLMs on the market to be covered by risk-mitigation obligations, such as the ones already in place in the DSA. Within the DSA, these include an obligation to designate a point of contact for the recipients of the service,[114] transparency obligations to publish at least once a year a report on their content moderation practices and the complaints received,[115] and obligations to put mechanisms in place to allow users to notify them of the presence of information that may be illegal.[116] And if the user base is significant, providers

---

[110] Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges Competencies, and Strategies*, 29 HARV. J. L. & TECH. 353, 356 (2016)

[111] Kaminski, *supra* note 31, at 3; *see also* National Institute of Standards and Technology, *AI Risk Management Framework*, https://perma.cc/QL9G-97RE; N.Y.C. ADMIN. CODE § 20-871 (2023); Efroni Zohar, *The Digital Services Act: Risk-Based Regulation of Online Platforms*, INTERNET POLICY REVIEW OPINION (Nov. 16, 2021), https://perma.cc/7YJ9-PXVZ.

[112] Kaminski, *supra* note 31, at 4, 8.

[113] Along these lines, the EU is also discussing a Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM (2022) 496 final (Sept. 28, 2022), 2022/0303 (COD). For a commentary and critique, see Wendehorst, *supra* note 31.

[114] DSA, *supra* note 19, Article 12.

[115] *Id.* Article 15.

[116] *Id.* Article 16.

could be required by the DSA to "analyze and assess any systemic risk stemming from the design or functioning of their service and its related systems" to the information environment, such as the proliferation of hate speech and harmful or dangerous content.[117]

They could then be obliged to take measures to mitigate those identified risks. These measures could have a variety of forms: Flagging or water-marking AI-generated content to help identify whether a piece of text was written by AI, for instance, could be especially useful if the text includes inaccuracies or is about a sensitive topic, such as politics.[118] Providers could also be obliged to have certain mandatory safeguards that train algorithms to not generate foreseeably harmful content. Providers could also be obliged to take measures that prevent workarounds of moderation tools (for instance, if systems are trained not to answer questions about a particular topic but still answer when asked to write a poem about that topic).[119]

This, of course, would not be a silver bullet. Some of the obligations included in the DSA may make less sense for LLMs—such as the obligation to provide a statement of reasons for any restrictions on visibility of user-generated content (such as a tweet or a post on Facebook).[120] Similarly, there may be obligations that are not in the DSA that should be imposed on the providers of LLMs, such as more transparency of the training data and the capabilities of these models, or defining certain instances where their use should be prohibited.[121] If these obligations are accompanied by a safe harbor for LLM output that complies with the obligations,

---

[117] *Id.* Article 34.

[118] Watermarking is a technology that inserts special words or content into LLM generated text, that makes it easier to detect later *see* Keith Collins, *How ChatGPT Could Embed a 'Watermark' in the Text It Generates*, N.Y. TIMES (Feb. 17, 2023).

[119] *See* Jack Cushman, *ChatGPT: Poems and Secrets*, LIBRARY INNOVATION LAB (Dec. 20, 2022), https://perma.cc/SPJ5-992U.

[120] *See* DSA, *supra* note 19, Article 17.

[121] The May 2023 version of the AI Act includes transparency obligations for providers of foundation models like LLMs *see* European Parliament, *AI Act: A Step Closer to the First Rules on Artificial Intelligence*, NEWS-EUROPEAN PARLIAMENT (May 11, 2023), https://perma.cc/3SU7-U4SY; for a discussion on whether certain uses should be banned *see*, *e.g.*, Julia Landwehr, *People Are Using ChatGPT in Place of Therapy—What Do Mental Health Experts Think?,* HEALTH.COM (May 13, 2023), https://perma.cc/R77A-NYHB.

then these fast-growing and developing technologies will have room for innovation, while still being subject to necessary democratic and societal controls.

III. LLMs IN THE LIGHT OF THE DSA: AN ALTERNATIVE INTERPRETATION OR A REGULATORY PROPOSAL

So far, we have established that LLMs raise challenges and risks to the information environment that are similar to those raised by social media. We also established that, in general, pre-DSA speech regulation and liability law has not seemed adequate to address them, and that, in Europe, the DSA introduced an interesting risk-regulatory framework to address these risks, mostly focused on social media. We ask now: Does the DSA also apply to LLM providers such as ChatGPT?

This section argues that when LLMs can be reasonably considered analogous to a search engine, it does. Indeed, the DSA applies to search engines as well. Though OpenAI's CEO, Sam Altman, has cautioned that people shouldn't be relying on ChatGPT "for anything important," and Google claims that Bard is not a search engine but "a complement to search,"[122] many users are nonetheless using LLMs as search engines.[123] As David Pierce wrote for The Verge,

> [T]he thing about Bard—and really the thing about every chatbot including ChatGPT and the new Bing—is that Google doesn't actually get to choose how you use it. People have spent the last few months using ChatGPT to replace a search engine.[124]

This section thus argues that if LLMs are being placed on the market in such a way that they can reasonably be used as search engines, they should be subject to the regulation applied to search engines. Under the DSA, this means they would both benefit from the safe harbor and be bound by substantive obligations such publishing a yearly transparency report. In addition, if their user base in Europe is considerable, they would be bound by the DSA's risk mitigation obligations.

I develop this argument in three steps: First, I briefly explain why some people argue that the DSA does not apply to services like ChatGPT. Basically, these services are very different from social media platforms: they don't host user-generated content, but rather generate it themselves after being prompted. Second, I briefly

---

[122] Pierce, *supra* note 34.

[123] *See, e.g.,* Will Knight, *The Race to Build a ChatGPT Powered Search Engine*, WIRED (Feb. 6, 2023), https://perma.cc/AB6R-BA9Y; Nico Grant & Cade Metz, *A New Chat Bot Is a 'Code Red' for Google's Search Business*, N.Y. TIMES (Dec. 21, 2022).

[124] Pierce, *supra* note 34.

discuss the complicated history of search engine regulation. Though they are also not traditional online intermediaries, one of the objectives of the DSA was to cover new intermediary technologies, like search engines. Third, I explain how and why DSA's legal regulation of search engines should be extended to LLMs, especially when they are being placed in the market in a way that allows users to use them as search engines. But not all of the measures of the DSA make sense for LLMs, and thus I finish by clarifying what these limits are and making some preliminary proposals for policymakers to consider.

A. ChatGPT and the Like Are Not Traditional "Hosting" Services

Recall that the DSA applies to "intermediary services offered to recipients of the service that have their place of establishment or are located in the Union."[125] The Act defines "intermediary services" as "one of the following information society services: (i) a 'mere conduit' service . . . (ii) a 'caching' service . . . or (iii) a 'hosting' service."[126] Mere conduit and caching services consist of the technical transmission of information and the intermediate and temporary storage of that information for the purpose of making the transmission more efficient.[127] These are mainly internet service providers or direct messaging services. Hosting services consist of "the storage of information provided by, and the request of, a recipient of the service."[128] Social media, for example, involves hosting user-generated content commonly known as posts (or tweets). As explained above, the DSA creates a safe harbor from liability that these services could face for carrying illegal content, subject to certain due diligence conditions, and it creates additional risk-mitigation obligations for "very large online platforms" and "very large online search engines."[129]

At first sight, services like ChatGPT are not covered by the DSA: They neither consist of the merely technical transmission of information nor host user generated content. Rather, they host AI generated content. Philipp Hacker, Andreas Engel, and Theresa List argue that even if LLMs create concerns about loss of trustworthiness and the deterioration of the information environment, risks with which

---

[125] DSA, *supra* note 19, Article 2.

[126] *Id.* Article 3(g).

[127] *Id.* Article 3(g)(i)–(ii).

[128] *Id.* Article 3(g)(iii).

[129] *Id.* Articles 34, 35, 36.

researchers and policymakers are already familiar, the DSA is not fit for dealing with those risks:

> [T]he DSA only applies to so-called intermediary services (Article 2(1) and (2) DSA). These are conclusively defined in Article 3(g) DSA. They cover Internet access providers, caching services, and "hosting" services such as social media platforms. LLMs, arguably, do not qualify as either of these instances. Hosting services come closest, but they require the storage of information provided by, and at the request of, a user (Article 3(g)(iii) DSA). The trick with LLMs, however, is that the relevant content is decidedly not provided by the user, but by the LLM itself, having been prompted by the user via the insertion of certain query terms (e.g., "write an essay about content moderation in EU law in a lawyerly style"). With the DSA arguably inapplicable, the regulation of LLM content is left to the thicket of Member State speech regulation, which not only varies considerably across the EU, but also generally lacks the DSA instruments aiming to guarantee the speedy removal of harmful speech from the online sphere.
>
> Of course, the DSA will apply if a user posts AI-generated content on a social network, such as Twitter. But at this point, it is often already too late to stem the tide of disinformation, hate speech, or manipulation. With LLMs, it is the creation that matters.[130]

This analysis, however, does not consider that the DSA applies to search engines, which are intermediaries that also do not fit easily within the definition of hosting services. Indeed, as discussed below, search engines also don't host user-generated content. What follows explains why search engines are covered by the DSA and why some LLM-powered services should be legally considered search engines.

B. The Legal Definition of a Search Engine

The question of how search engines fit into the EU's intermediary liability regulation has a long history in Europe.[131] Broadly speaking, search engines help end-users find and effectively retrieve information that is publicly available on the Internet. As a legal category, however, search engines were not defined in the E-Commerce Directive.[132] They also don't fit squarely in the definition of any of the

---

[130] *See* Hacker, Engel & List, *supra* note 27; *see also* Hacker, Engel & Mauer, *supra* note 17.

[131] *See* Joris van Hoboken, *Search Engine Freedom: On the Implications of the Right to Freedom of Expression for the Legal Governance of Web Search Engines* (Ph.D. thesis, University of Amsterdam 2012), https://perma.cc/L542-FT4D.

[132] DSA, *supra* note 19, Recital 28.

intermediaries covered by the E-Commerce Directive or the DSA: they do not just transmit information nor do they host "information provided by a recipient of the service."[133] Search engines offer hyperlinks based on material that their algorithm finds using their indexes in response to a user search query.[134]

This lack of clarity led to fragmentation in national legislation and jurisprudence on liability regimes for search engines.[135] They were considered a mere conduit activity in the UK and a form of hosting in Germany, while Spain and Portugal simply extended the liability exemption by law to search engine activities.[136] In some countries, the fact that certain intermediaries not only stored and made accessible user-generated materials but also organized them, indexed them, linked to ads, and so on, was interpreted as too active and, therefore, not falling under the protection of the E-Commerce Directive.[137]

The Court of Justice never explicitly addressed whether search engines were hosting services under the E-Commerce Directive, but settled part of this question in *Google Search* (2010) where it held that the AdWords-referencing services was a hosting service.[138] In this case, one of the questions was whether Google AdWords could be held liable for showing trademark-infringing content.[139] The question

---

[133] *See, e.g.,* Van Hoboken, *supra* note 131.

[134] *See* Giovani Sartor, *Providers Liability: From the eCommerce Directive to the Future* 25 (European Parliament 2017), https://perma.cc/Y94C-7PF5.

[135] Nevertheless, when many Member States implemented the E-Commerce Directive they extended hosting liability to search engines. Tambiama Madiega, *Reform of the EU liability Regime for Online Intermediaries: Background on the Forthcoming Digital Services Act*, European Parliamentary Service PE 649.404, May 2020, at 8, https://perma.cc/UF3B-LXWA; Van Hoboken, *supra* note 131, at 217.

[136] *Id*. at 4.

[137] Sartor, *supra* note 134, at 25.

[138] AdWords is a paid referencing service that "enables any economic operator, by means of the reservation of one or more keywords, to obtain the placing, in the event of a correspondence between one or more of those words and that/those entered as a request in the search engine by an internet user, of an advertising link to its site. That advertising link appears under the heading 'sponsored links,' which is displayed either on the right-hand side of the screen, to the right of the natural results, or on the upper part of the screen, above the natural results." Joined Cases C-236/08 & C-237/08, *Google France SARL v. Louis Vuitton Malletier SA,* [2010] ECR I-2417, at 23 (Mar. 23, 2010) [hereinafter *Google Search*].

[139] *Id*.; *see also* Van Hoboken et al., *supra* note 58, at 12.

thus was whether an internet referencing service constituted a hosting information society service within the meaning of the E-Commerce Directive.[140] The Court considered that Google was transmitting information from the advertisers to other users and storing "on its server, certain data, such as the keywords selected by the advertiser, the advertising link and the accompanying commercial message, as well as the address of the advertiser's site."[141] The Court explained that, in order to fit the definition of a hosting service, it is necessary that the provider plays a neutral role, in the sense that it should be "merely technical, automatic and passive," and with a "lack of knowledge or control of the data which it stores."[142]

Applying this reasoning to AdWords, the Court first explained that data processing and "the resulting display of the ads is made under conditions which Google controls."[143] Thus, it seemed like Google does not fall under the definition of hosting. However, the Court did clarify that the mere facts that (1) Google sets the service terms and provides general information to its clients, and that (2) there is "concordance between the keyword selected and the search term entered by an internet user" were not themselves sufficient to conclude that Google had control over the data entered in the system and stored in its service.[144] This led the Court to establish that sponsored links services could be considered hosting services.[145]

In subsequent years, different studies commissioned by the European Parliament in anticipation of the DSA recommended that the legal notions that underpin the safe harbor regime be clarified and, specifically, that search engines should benefit from the safe harbor.[146] The DSA thus sought to solve some of this ambiguity and defined an online search engine in Article 3(i) as

---

[140] *Id.* at 106.

[141] *Google Search*, *supra* note 138, at 111, 112.

[142] *Id.* at 116, 117.

[143] *Id.* at 115.

[144] *Id.* at 116, 117.

[145] *See* Paul Przemysław Polanski, *Technical, Automatic and Passive: Liability of Search Engines,* 6 J. INT'L COM. L. & TECH. 1, 49 (2011) (arguing also that the ruling widened the definition of hosting services, and that "a liberal application of hosting exemption to all instances of websites storing third-party content may cause problems").

[146] Madiega, *supra* note 135, at 17; Sartor, *supra* note 134 (arguing that that the exemption should be explicitly extended to ensure the provision of these services).

> an intermediary service that allows users to input queries in order to perform searches of, in principle, all websites, or all websites in a particular language, on the basis of a query on any subject in the form of a keyword, voice request, phrase or other input, and returns results in any format in which information related to the requested content can be found.[147]

The DSA, however, copied the exact same definition of hosting, mere conduit, and caching services from the E-Commerce Directive and it extended the liability to each of these services specifically, but not generally. [148] Nowhere did it clarify explicitly what search engines are.

Under a strict, formalistic interpretation of the DSA, this is perhaps problematic: recall that, in principle, the application of the safe harbor depends on whether (a) a service is an information society service ("provided for remuneration, at a distance, by electronic means and at the individual request of a recipient"[149]) and (b) the providers' activity falls into one of those three categories.[150] Though sponsored links services are hosting services following *Google Search,* the case is slightly less clear for search engines. With sponsored links services, advertisers request and pay Google to store and transmit their information to users who enter a matching keyword. In the case of organic search engine results, some commentators have noted that it is less clear that the autonomous indexing of websites is done at the request of online publishers.[151]

This interpretation of the DSA would be wrong. The DSA is very clear that search engines are covered by it. First, the DSA extends all the substantive and risk-

---

[147] DSA, *supra* note 19, Article 3(J).

[148] See for example, the wording of Article 6(1): "Where an information society service is provided that consists of the storage of information provided by a recipient of the service, the service provider shall not be liable for the information stored at the request of a recipient of the service." DSA, *supra* note 19, Article 6(1).

[149] *See supra* note 60 and accompanying text on information society services.

[150] *See* Section I.B.1.

[151] See Sartor, *supra* note 134, at 25, proposing two arguments in favor of and against concluding that search engines are hosting services: "No, they do not provide hosting, since they autonomously index web-sites and determine the outcomes of searches. Yes, they do provide hosting, since they are implicitly authorized by online publishers (uploaders) to index all content on the open web, and make it accessible through an algorithm meant to satisfy user's preferences."

mitigation obligations specifically to very large search engines.[152] Second, the DSA's Recitals explain that the E-Commerce Directive was updated in part because the online ecosystem was significantly more complex than in 2000.[153] The DSA was supposed to be future-proof and new providers of internet intermediary services should be able to benefit from exemptions from liability when they facilitate and support the functioning of the internet by, for instance, aiding users in finding information. In this sense, Recital 28 specifically included search engines:

> Since 2000, new technologies have emerged that improve the availability, efficiency, speed, reliability, capacity and security of systems for the transmission, 'findability' and storage of data online, leading to an increasingly complex online ecosystem. . . Such services include . . . online search engines, cloud infrastructure services, or content delivery networks, that enable, locate or improve the functions of other providers of intermediary services. Those services, too, can benefit from the exemptions from liability to the extent that they qualify as 'mere conduit', 'caching' or 'hosting' services.[154]

EU law has evolved to include search engines in its intermediary liability regime, and they are covered by the DSA. The next section explains why similar reasoning should be applied to LLMs that function like search engines.

C. Certain LLMs Should Be Considered Search Engines by Analogy

Since search engines are clearly covered by the DSA, by analogy certain LLMs are covered by the DSA as well. There are two reasons that support this interpretation:

First, like search engines, LLM applications like ChatGPT do not squarely fit within a strict definition of hosting, mere conduits, and caching services. They, however, could fit into the broader definition of a hosting service adopted by the Court in *Google Search* and recent caselaw. Tools like ChatGPT and Bard usually store query and user information in their servers, like Google AdWords does. They also play a role that is neutral, in the sense that such an AI program has "no know-

---

[152] See DSA, *supra* note 19, Section 5 on "Additional obligations for providers of very large online platforms and of very large online search engines to manage systemic risks."

[153] Recitals are non-binding provisions at the beginning of EU Acts that are considered important in interpreting ambiguous provisions. *Recitals*, THOMSON REUTERS PRACTICAL LAW, https://uk.practicallaw.thomsonreuters.com/w-009-6368?contextData=(sc.Default)&transitionType=Default&firstPage=true.

[154] *See* DSA, *supra* note 19, Recital 28.

ledge and control over the content it stores"[155] or the content it generates, because the content is generated through means that are technical, automatic, and thus passive. [156] Indeed, in a recent case involving YouTube's content moderation practices, the Court explained that algorithmic content moderation—the fact that a platform "implements technological measures aimed at detecting . . . content . . . does not mean that, by doing so, that operator plays an active role giving it knowledge of and control over the content."[157] ChatGPT's answers are assembled using algorithms that predict what makes sense as the next word, based on a user's prompt. [158] OpenAI, thus, does not have any knowledge or an active role in controlling the content ChatGPT generates. It could thus be argued that its role is neutral in a similar way to how YouTube is neutral in hosting user-generated content.

Unlike Google AdWords and YouTube, however, a user's interaction with an LLM like ChatGPT does not directly involve a third party who also requests the service. There is no transmission of user data from an advertiser to a consumer, or from a social media user to others. This is, however, where the functional analogy with search engines is key. When a user conducts a search on Google, there is also no clear third party. The results Google shows me are generated for me alone, upon my request alone. Additionally, in the case of established search engines, a great deal of prediction and analytics goes into delivering those results. Search engines must facilitate access to all online material, while providing answers to specific queries that are valuable to a particular user. As Joris van Hoboken explains, "this type of intelligent guessing is precisely what offering a search engine is all about: to select and rank a list of online resources that has a good—or better, as high as possible—chance of satisfying the demand of the user as imperfectly expressed in a search query."[156] That sounds like ChatGPT too.

Second, the ECJ has stated that intermediary liability rules in the DSA must be interpreted not only in the light of their wording, "but also of its context and the

---

[155] Joined Cases C-682/18 & C-683/18, Frank Peterson v. Google LLC, and others and Elsevier Inc. Cyando AG, 2021 ECJ EUR-Lex Lexis 503, at 106 (June. 22, 2021) [hereinafter *YouTube*].

[156] *Google Search*, *supra* note 138, at 113; see *supra* Section III.B, on how storing this kind of information was part of what led the Court to determine that Google AdWords was a hosting service.

[157] *YouTube*, *supra* note 155, at 109.

[158] OpenAI, *supra* note 2, at 1.

[156] Van Hoboken, *supra* note 131, at 51.

objectives of the legislation of which it forms part."[159] Paying special attention to the context and the objectives of the DSA, it is clear that some LLM applications allow users to perform searches of information on the whole world wide web (or at least some snapshot of a significant part of the web), just like search engines. ChatGPT and the like invite users to ask something such as "Explain quantum computing in simple terms."[160] Additionally, its answers are drawn from having been trained with an immense data set of different internet sources ranging from web pages to books and research articles.[161]

LLMs, like ChatGPT, do not provide links but rather what seems like summarized content (often full of inaccuracies) or a text box with references. This is different from search engines. Functionally, however, that is a key part of what search engines do for users. As Andrei Broder noted, search engines help us address three main needs: (i) navigating to specific online locations, (ii) finding information about specific topics, and (iii) finding purchasing opportunities, services, and online resources.[162] And the DSA's definition of search engine is agnostic as to the form of the results: It characterizes search engines as "return[ing] results in any format in which information related to the requested content be found."[163]

Google has been using AI models to summarize search results for users.[164] This task of "organization" is by no means easy, and by no means neutral. It forces old and new search engines and online intermediaries to engage in content moderation, and has long raised questions.[165] Even if today's LLMs are still often flawed, LLM-powered tools like Bard, ChatGPT, and Bing may be yet another iteration of what search engines and platforms have long been trying to do: organize information in our information-rich environment.

---

[159] *Google Search, supra* note 138, at 48.

[160] *See* ChatGPT, https://chat.openai.com/chat.

[161] OpenAI, *Technical Report*, *supra* note 2, at 42, 53.

[162] Andrei Broder, *A Taxonomy of Web Search*, 36 ACM SIGIR FORUM, 2 (2002) (quoted by Van Hoboken, *supra* note 131, at 52).

[163] DSA, *supra* note 19, Article 3(j).

[164] Pierce, *supra* note 34.

[165] *See, e.g.*, Kari Paul, *Google Misdirects One in 10 Searches for Abortion to 'Pregnancy Crisis Centers*,' GUARDIAN (June 9, 2022), https://perma.cc/F4DJ-26VZ; Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos (AEPD), Mario Costeja González, Case C-131/12, [2014] QB 1022 (ECJ 2014).

D. The Advantages and the Limits of this Proposal

To summarize, this Essay is proposing that LLMs be regulated like search engines and large tech platforms in the EU, especially in instances when these tools are being placed on the market in a way that can be reasonably assumed to be used for search purposes. In the EU this can stem from a functional interpretation of the DSA that examines the regulation of the service in context and adopts a broader interpretation of hosting services, like the Court has done before in previous cases. This would be the case when these tools let the public submit queries about the world (rather than, for example, being plugged into workflow manager platforms to take notes or summarize meetings).

There are several limitations to this proposal, and it is therefore offered as a partial measure that could be supplemented with more suitable frameworks when needed. This leaves out many applications that are also designed to produce content, such as AI powered chatbots that may be designed to provide guidance in specialized domains, like law or medicine, where other special requirements may arise.[166]

However, this legal interpretation, or intervention if adopted elsewhere, has two important advantages in general, and one particular to the European context:

First, the LLMs and other chatbots would benefit from the safe harbor if they remove harmful content when they are made aware of it.[167] Despite their risks, these tools have great potential to assist individuals and firms in a variety of tasks. Developers could benefit from legal certainty as to whether, when, and to what extent they could be liable for illegal content generated by the general use of LLM-powered chatbots.

Second, these tools sometimes create significant risks, and it would be wise to require the developers of these tools to run impact assessments, evaluate the potential impact of these tools on different foreseeable societal risks, and adopt mitigation measures. So far, we see OpenAI doing some of this internal auditing

---

[166] See Bambauer, *supra* note 1, at 355, discussing this under US law.

[167] *See* Eugene Volokh, *Large Libel Models? Liability for AI Outputs*, 3 J. FREE SPEECH L. 489, 514–18 (2023) (explaining that this would in practice require providers to take reasonable steps to block certain outputs once they have been made aware that their system is generating a form of illegal or harmful speech).

already,[168] but it is unclear whether they are obliged to do so, and there are no clear legal mechanisms that will guarantee the auditing of their work. If the proposal here were to be adopted, companies placing these tools on the market would be required to, among other things, set in place easily accessed mechanisms that let any individual or entity notify them of the presence on their service of specific items of information that the notifier considers to be illegal content.[169] This would trigger the blocking obligations which would also help the AI companies improve their own systems. The proposal would also require the companies to make publicly available reports on their content moderation practices.[170] In certain instances, for example if the amount of illegal output is considerable, the AI companies could be obliged to conduct yearly systemic risk assessments,[171] and would have to put in place mitigation measures to address the identified systemic risks.[172]

Third, the DSA is already in effect.[173] Applying the DSA will be a useful way to create legal certainty for companies since they will benefit from the safe harbor. At the same time, companies would be bound by risk-mitigation obligations right away. In Europe, the AI Act will most likely include specific obligations for LLMs. But it will be finalized by the end of 2023, and there will likely be a transition period of 2 to 3 years before it is fully applicable.[174] That is too long from now.

Conclusion

Like other AI systems and social media, LLMs have been found to present false or invented information confidently, to facilitate access to what can be dangerous information, and to reproduce certain forms of social bias in its answers.[175] At the same time, LLMs can be used in a wide range of applications such as dialogue

---

[168] See OpenAI, *Technical Report, supra* note 2, for a rather thorough technical report on the risks identified by OpenAI and the mitigation efforts they have undertaken.

[169] DSA, *supra* note 19, Article 16.

[170] *Id.* Article 15.

[171] *Id.* Article 34.

[172] *Id.* Article 35.

[173] *See* European Parliament, *supra* note 59.

[174] Edwards, *supra* note 25.

[175] *See* Section II; OpenAI, *Technical Report, supra* note 2, at 47.

systems, text summarization, and machine translation; indeed, they are already being deployed to do that.[176]

This Essay proposes that policymakers should aim to sustain and encourage innovation by protecting these systems' developers from excessive liability, while imposing some due-diligence risk-mitigation obligations to address the risks these models raise in relation to the degradation of the information environment. This is not aimed at precluding future, more specific, regulations. A model for the system this Essay proposes is the European Digital Services Act, which extends broad liability exemptions for online intermediaries when certain conditions are met, while imposing duties to conduct risk assessments and adopt mitigation obligations for the largest actors. And in the EU, lawmakers, and adjudicators could apply the DSA directly to general purpose LLMs that, like ChatGPT or Bard, are being used as search tools.

---

[176] *See, e.g.*, Simon Torkington, *How Might Generative AI Change Creative Jobs*, WORLD ECONOMIC FORUM (May 9, 2023), https://perma.cc/3BUD-A929 (discussing how the creative and marketing industries are adapting and adopting generative AI).