

Online Disinhibition Effect: Tinjauan Sistematis dan Implikasinya terhadap Etika Kewargaan Digital

Habibah Pidi Rohmatu^{a,1*}, Raistin Nur Abidin^{b,2}, Alex^{c,3}, Yonas Belay Gela^{c,4}

^{a,c}Pendidikan Pancasila dan Kewarganegaraan, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Lambung Mangkurat

^bPendidikan Pancasila dan Kewarganegaraan, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Pamulang

^cDepartment of Civics and Ethical Studies, Dilla University

¹habibahpidi@ulm.ac.id ; ²alexkip@ulm.ac.id ; ³dosen02860@unpam.ac.id ; ⁴yonasbelay118@gmail.com

*korespondensi penulis

Naskah diterima: 3 Agustus 2026, direvisi: 30 Agustus 2026, disetujui: 25 September 2026

Abstrak

Perkembangan teknologi digital telah mengubah pola interaksi manusia dan memunculkan fenomena *Online Disinhibition Effect* (ODE), yang berimplikasi pada perilaku dan etika dalam ruang digital. Namun, kajian mengenai *Online Disinhibition Effect* (ODE) masih menunjukkan fragmentasi konseptual dan metodologis, yang ditandai oleh dominasi pendekatan kuantitatif berbasis *cross-sectional*, keterbatasan eksplorasi konteks sosial-budaya, serta kecenderungan fokus pada dimensi *toxic disinhibition* dibandingkan aspek prososialnya, sehingga menghasilkan pemahaman yang parsial terhadap kompleksitas fenomena ini. Penelitian ini bertujuan untuk mensintesis fenomena ODE, faktor yang memengaruhinya, dampaknya terhadap etika digital, serta implikasinya dalam perspektif etika kewargaan digital. Penelitian menggunakan pendekatan *Systematic Literature Review* (SLR) dengan protokol PRISMA terhadap artikel dari Scopus, Web of Science, dan Google Scholar. Hasil penelitian menunjukkan bahwa ODE dipengaruhi oleh interaksi antara faktor teknis platform (anonimitas, invisibilitas, asinkronitas), faktor psikologis (kontrol diri, empati, *moral disengagement*), dan faktor sosial-kontekstual. Dampak ODE didominasi oleh *toxic disinhibition*, seperti *cyberbullying*, *trolling*, penyebaran hoaks, serta pelanggaran norma komunikasi, yang berimplikasi pada degradasi kualitas ruang publik digital. Penelitian menegaskan bahwa ODE merupakan fenomena sosio-teknis yang memerlukan pendekatan integratif. Oleh karena itu, disarankan penguatan pendidikan etika kewargaan digital, pengembangan regulasi diri pengguna, serta desain platform yang meningkatkan akuntabilitas dan meminimalkan perilaku toksik.

Kata-kata kunci: Online Disinhibition Effect; Perilaku Daring; Etika Kewargaan Digital

Abstract

The rapid development of digital technology has transformed patterns of human interaction and given rise to the Online Disinhibition Effect (ODE), which has significant implications for behavior and ethics in digital spaces. However, existing studies on ODE exhibit conceptual and methodological fragmentation, characterized by the dominance of cross-sectional quantitative approaches, limited exploration of socio-cultural contexts, and a tendency to focus on toxic disinhibition rather than its prosocial dimensions, resulting in a partial understanding of this complex phenomenon. This study aims to synthesize the phenomenon of ODE, its influencing factors, its impacts on digital ethics, and its implications from a digital citizenship ethics perspective. This research employs a Systematic Literature Review (SLR) using the PRISMA protocol, analyzing selected articles from Scopus, Web of Science, and Google Scholar. The findings reveal that ODE is shaped by the interaction between platform affordances (anonymity, invisibility, asynchronicity), individual psychological factors (self-control, empathy, moral disengagement), and socio-contextual dynamics. The impacts of ODE are predominantly associated with toxic disinhibition, including cyberbullying, trolling, misinformation, and violations of communication norms, which contribute to the degradation of digital public discourse. The study concludes that ODE is a socio-technical phenomenon requiring an integrative approach. It is therefore recommended to strengthen digital citizenship education, enhance users' self-regulation capacities, and promote platform designs that improve accountability and reduce toxic behavior.

Keywords: Online Disinhibition Effect; Online Behaviour; Digital Ethics Citizenship

Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah mengubah cara manusia berinteraksi, menciptakan ruang digital yang memungkinkan komunikasi tanpa batas geografis dan temporal. Dalam kondisi ini, ruang digital memunculkan fenomena psikologis yang khas, salah satunya adalah *Online Disinhibition Effect* (ODE). Konsep ODE pertama kali diperkenalkan oleh John Suler (2004), yang menjelaskan bahwa individu cenderung menunjukkan perilaku yang berbeda secara daring dibandingkan dengan interaksi tatap muka. Perbedaan tersebut dapat muncul dalam bentuk *benign disinhibition*, yaitu peningkatan keterbukaan dan ekspresi diri secara positif, maupun *toxic disinhibition*, yang ditandai dengan munculnya perilaku agresif, tidak etis, dan melanggar norma sosial.

Dalam konteks etika penggunaan internet di Indonesia, ODE menjadi perhatian serius karena berkontribusi terhadap berbagai bentuk perilaku tidak etis, seperti *trolling*, *cyber bullying*, agresi daring, dan ketidaksopanan siber (Febriana & Fajrianthi, 2019; Manalu, 2023; Matondang & Afdal, 2023; Rafly & Ningsih, 2025). Fenomena ini tidak dapat direduksi sebagai bentuk impulsivitas individu semata, melainkan hasil dari interaksi antara desain algoritma dan mekanisme psikososial pengguna.

Desain algoritma platform digital seperti anonimitas, invisibilitas, dan asinkronitas berpengaruh untuk mendistorsi persepsi akuntabilitas, menurunkan *self-awareness*, serta memfasilitasi *moral disengagement* dalam interaksi daring (Febriana & Fajrianthi, 2019; Manalu, 2023; Siregar, 2026; S. Wu et al., 2017). Dalam kerangka ini, ODE merepresentasikan bentuk *situated behavior*, di mana ruang digital

secara aktif membentuk dan memediasi ekspresi moral individu yang bekerja melalui algoritma yang tidak netral.

Implikasi dari kondisi tersebut melampaui ranah interpersonal dan memasuki dimensi ruang publik digital. Peningkatan kasus penyebaran hoaks, ujaran kebencian, serta praktik komunikasi yang tidak santun menunjukkan bahwa ODE berkontribusi terhadap degradasi kualitas deliberasi dalam ruang publik digital, yang pada akhirnya berpotensi mengganggu kohesi sosial dan praktik kewargaan (Gumelar et al., 2024; Matondang & Afdal, 2023; Sudarja, 2025).

Dalam perspektif etika kewargaan digital, kondisi ini menunjukkan adanya kesenjangan antara perkembangan teknologi digital dan internalisasi nilai-nilai etis. Meskipun faktor seperti kontrol diri dan regulasi emosi diidentifikasi sebagai variabel protektif terhadap ODE, studi menunjukkan bahwa aspek tersebut belum secara sistematis diintegrasikan dalam kerangka pendidikan etika digital (Bulan & Wulandari, 2021; Khairani & Guspa, 2025).

Sementara itu, penelitian mengenai ODE menunjukkan keterbatasan metodologis yang berimplikasi pada terbatasnya generalisasi temuan. Pertama, dominasi penggunaan sampel mahasiswa dalam berbagai studi menyebabkan representasi populasi menjadi sempit dan belum mencerminkan keragaman pengalaman pengguna digital secara lebih luas. Kedua, penggunaan desain *cross-sectional* dan data *self-report* membatasi kemampuan penelitian dalam menjelaskan hubungan kausal, serta berpotensi menimbulkan bias persepsi dan pelaporan. Ketiga, masih terbatasnya eksplorasi terhadap faktor budaya dan konteks sosial mengakibatkan pemahaman mengenai ODE cenderung bersifat homogen dan kurang sensitif terhadap perbedaan lingkungan sosial. Keempat, variasi dalam

instrumen pengukuran yang digunakan antar studi menyulitkan perbandingan dan sintesis hasil penelitian secara konsisten lintas konteks. Kelima, integrasi kerangka konseptual masih relatif lemah, dengan kecenderungan penelitian yang lebih menitikberatkan pada *toxic disinhibition* dibandingkan eksplorasi dimensi *benign disinhibition*. (Bulan & Wulandari, 2021; Cheung et al., 2020; Fakhri, 2025; Febriana & Fajrianthi, 2019; Gan et al., 2024; Haqie et al., 2024; Juliati et al., 2025; Khairani & Guspa, 2025; Lestari & Alwi, 2024; Manalu, 2023; Matondang & Afdal, 2023; Rafly & Ningsih, 2025; Shabira & Lestari, 2024; Siregar, 2026; Soares et al., 2023; C. Wang et al., 2014; L. Wang et al., 2024; B. Wu et al., 2022).

Meskipun penelitian mengenai *Online Disinhibition Effect* (ODE) terus berkembang, masih terdapat beberapa *gap* yang perlu diperhatikan. Pertama, penelitian mengenai ODE masih tersebar dalam berbagai konteks, konstruk, dan pendekatan analisis. Studi terdahulu telah mengidentifikasi faktor teknis, psikologis, dan sosial yang berkaitan dengan ODE serta berbagai manifestasinya, tetapi hubungan antara faktor-faktor tersebut, bentuk *benign* dan *toxic disinhibition*, serta konsekuensi perilakunya belum banyak disintesis dalam satu kerangka yang komprehensif. Kedua, khususnya dalam keseimbangan kajian antara *toxic* dan *benign disinhibition*. Literatur cenderung memberikan perhatian lebih besar terhadap perilaku agresif dan tidak etis, sementara potensi konstruktif ODE dalam bentuk *self-disclosure*, keterbukaan emosional, dukungan sosial, dan partisipasi masih relatif kurang dieksplorasi. Ketiga, terdapat *gap* metodologis dan kontekstual, yang terlihat dari dominasi sampel mahasiswa, desain *cross-sectional*, penggunaan data *self-report*, keterbatasan studi lintas budaya dan lintas platform, serta variasi instrumen pengukuran

yang menyulitkan komparasi temuan. Keempat, terdapat gap pada pemaknaan ODE dalam kajian kewargaan digital, karena sebagian besar penelitian masih menempatkan ODE sebagai persoalan psikologis dan perilaku pengguna, sementara implikasinya terhadap tanggung jawab dan akuntabilitas warga dalam ruang digital, norma komunikasi, kualitas ruang publik digital, dan pendidikan etika kewargaan belum banyak diintegrasikan secara sistematis. Keempat kesenjangan tersebut menunjukkan bahwa diperlukan suatu tinjauan sistematis yang juga mensintesis faktor pembentuk, dimensi *benign* dan *toxic*, dampak terhadap perilaku digital, serta keterbatasan penelitian untuk memperoleh pemahaman yang lebih utuh mengenai posisi ODE dalam pembentukan etika kewargaan digital.

Meskipun jumlah penelitian terus berkembang, kajian yang secara sistematis mengintegrasikan temuan-temuan tersebut dalam satu kerangka analisis yang komprehensif, terutama yang mengaitkannya dengan konteks pendidikan dan etika kewargaan digital, masih terbatas. Oleh karena itu, penelitian ini menggunakan pendekatan *systematic literature review* untuk mensintesis secara menyeluruh penelitian mengenai ODE, mengidentifikasi kecenderungan dan keterbatasan yang ada, serta menganalisis implikasinya terhadap penguatan etika kewargaan digital, sehingga dapat memberikan landasan konseptual yang lebih terstruktur bagi pengembangan penelitian dan praktik pendidikan di era digital.

Oleh sebab itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan melakukan *Systematic Literature Review* (SLR) terhadap penelitian mengenai *Online Disinhibition Effect* (ODE) melalui lima fokus analisis yang saling berkaitan. Pertama, penelitian memetakan fenomena dan faktor yang membentuk ODE

dalam interaksi di ruang digital, termasuk interaksi antara karakteristik teknis platform, faktor psikologis individu, dan kondisi sosial-kontekstual. Kedua, penelitian mensintesis dimensi *benign* dan *toxic disinhibition* untuk memperoleh pemahaman yang lebih seimbang mengenai manifestasi ODE. Ketiga, penelitian menganalisis dampak ODE terhadap perilaku dan etika digital, khususnya bentuk-bentuk perilaku yang berkaitan dengan kualitas interaksi dan ruang publik digital. Keempat, penelitian mengidentifikasi keterbatasan empiris, metodologis, konseptual, dan kontekstual dalam penelitian ODE yang dapat memengaruhi cakupan serta interpretasi temuan. Kelima, penelitian menafsirkan temuan tersebut dalam perspektif etika kewargaan digital untuk mengidentifikasi implikasinya terhadap tanggung jawab, akuntabilitas, norma komunikasi, dan penguatan pendidikan etika digital. Melalui kelima fokus tersebut, penelitian ini diharapkan menghasilkan sintesis yang lebih komprehensif mengenai ODE sekaligus memperluas pemaknaan fenomena tersebut dari persoalan perilaku individual menuju isu etika dan praktik kewargaan dalam ruang digital.

Metode

Penelitian ini menggunakan pendekatan *Systematic Literature Review* (SLR) dengan mengacu pada pedoman *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) 2020 (Page et al., 2021). Pemilihan PRISMA didasarkan pada karakteristik tujuan penelitian yang memerlukan sintesis terhadap temuan penelitian ODE yang tersebar dalam beragam konteks, populasi, konstruk, desain penelitian, dan pendekatan analisis. Dalam kondisi tersebut, proses identifikasi dan seleksi literatur perlu dilakukan melalui prosedur yang

eksplisit agar basis literatur yang digunakan dalam sintesis dapat ditelusuri dan dipertanggungjawabkan.

Sumber dan strategi pencarian literatur diperoleh dari tiga basis data ilmiah, yaitu Scopus, Web of Science, dan Google Scholar. Pencarian difokuskan pada artikel jurnal *peer-reviewed* yang secara eksplisit membahas *Online Disinhibition Effect* (ODE) dalam konteks interaksi dan perilaku digital. Untuk menghindari pembatasan pencarian pada salah satu dimensi ODE, strategi pencarian menggunakan istilah yang merepresentasikan konsep ODE secara umum serta kedua bentuk manifestasinya, yaitu: (“*online disinhibition effect*” OR “*online disinhibition*” OR “*benign disinhibition*” OR “*toxic disinhibition*”) AND (“*online interaction*” OR “*online behavior*” OR “*digital behavior*” OR “*cyberbullying*” OR “*trolling*” OR “*self-disclosure*”). Literatur dibatasi pada publikasi tahun 2015–2025 untuk menangkap perkembangan penelitian ODE dalam lingkungan digital kontemporer.

Hasil pencarian awal menghasilkan 206 artikel yang kemudian disaring untuk menghilangkan duplikasi dan artikel yang tidak relevan. Proses penyaringan dilakukan berdasarkan judul dan abstrak, diikuti dengan penilaian teks lengkap untuk memastikan kesesuaian dengan kriteria inklusi. Kriteria inklusi dalam penelitian ini meliputi: (1) artikel jurnal ilmiah yang telah melalui proses *peer-review*, (2) membahas *Online Disinhibition Effect* secara eksplisit, (3) berkaitan dengan perilaku atau interaksi digital, serta (4) memiliki relevansi dengan konteks pendidikan etika digital. Sementara itu, artikel yang bersifat opini, editorial, atau tidak memiliki keterkaitan langsung dengan ODE dikeluarkan dari analisis. Dengan demikian, diperoleh 35 artikel yang memenuhi kriteria inklusi awal. Sementara itu, artikel yang bersifat opini, editorial, atau tidak

memiliki keterkaitan langsung dengan ODE dikeluarkan dari analisis. Seluruh proses seleksi dilakukan secara sistematis dan berulang untuk memastikan konsistensi serta kesesuaian dengan kriteria yang telah ditetapkan.

Untuk meningkatkan validitas proses seleksi, setiap artikel ditelaah secara sistematis berdasarkan kriteria yang telah ditentukan. Artikel yang lolos tahap seleksi selanjutnya dianalisis menggunakan teknik analisis tematik untuk mengidentifikasi pola, faktor, dampak, serta keterbatasan penelitian terkait ODE. Proses analisis dilakukan dengan mengelompokkan temuan ke dalam kategori tematik yang sesuai dengan rumusan masalah penelitian, yaitu fenomena dan faktor ODE, dampak terhadap perilaku dan etika digital, keterbatasan penelitian, serta implikasinya terhadap etika kewarganegaraan digital.

Hasil dan Pembahasan

Fenomena dan Faktor *Online Disinhibition Effect* dalam Interaksi Ruang Digital

Penelitian ini menggunakan *Online Disinhibition Effect Theory* yang dikembangkan oleh John Suler, (2004) sebagai kerangka teoretis utama untuk memahami dan mensintesis fenomena *Online Disinhibition Effect* (ODE) dalam interaksi digital. Dalam teori ini, *disinhibition* merujuk pada kecenderungan individu untuk mengekspresikan diri secara lebih bebas, intens, dan tidak terkendali dalam interaksi daring dibandingkan dengan interaksi tatap muka. Lingkungan daring menghadirkan kondisi komunikasi yang berbeda melalui anonimitas, invisibilitas, asinkronitas, jarak psikologis, serta berkurangnya kehadiran sosial dan pengawasan langsung. Kondisi tersebut dapat mengubah cara individu mempersepsikan dirinya, lawan interaksi, serta konsekuensi sosial dari

tindakannya, sehingga ekspresi pikiran, emosi, dan perilaku dapat muncul dengan tingkat keterbukaan atau intensitas yang berbeda dibandingkan interaksi tatap muka.

Suler (2004) membedakan ODE ke dalam dua bentuk utama, yaitu *benign disinhibition* dan *toxic disinhibition*. *Benign disinhibition* merujuk pada peningkatan keterbukaan, ekspresi emosi positif, serta kecenderungan untuk berbagi pengalaman personal secara lebih jujur. Sebaliknya, *toxic disinhibition* ditandai oleh munculnya perilaku agresif, kasar, dan melanggar norma sosial, seperti ujaran kebencian, *cyberbullying*, dan perilaku antisosial lainnya. Meskipun kedua bentuk ini tampak dikotomis, dalam praktiknya keduanya seringkali saling berkelindan dan dipengaruhi oleh konteks serta kondisi psikologis individu.

Secara teoretis, ODE dijelaskan melalui enam mekanisme utama yang diidentifikasi oleh Suler, yaitu anonimitas disosiatif, invisibilitas, asinkronitas, introjeksi solipsistik, imajinasi disosiatif, dan minimisasi otoritas. Keenam mekanisme ini bekerja sebagai *psychosocial enablers* yang menurunkan hambatan normatif dalam interaksi daring. Namun, pada studi-studi mutakhir, tiga dimensi utama seperti anonimitas, invisibilitas, dan asinkronitas, muncul sebagai *core affordances* yang secara konsisten memediasi intensitas disinhibisi dalam ruang digital (Cheung et al., 2020).

Anonimitas, dalam konteks ini berfungsi sebagai mekanisme perlindungan identitas, sekaligus sebagai *disruptor of normative accountability*. Kemampuan untuk mengelola dan memodifikasi identitas daring (*identity controllability*) menciptakan kondisi di mana individu mengalami disosiasi antara identitas daring dan identitas sosialnya, sehingga menurunkan internalisasi konsekuensi moral atas tindakan yang dilakukan. Literatur menunjukkan

bahwa kondisi ini secara signifikan meningkatkan kecenderungan *toxic disinhibition*, terutama ketika tidak terdapat mekanisme regulasi eksternal seperti kepastian hukuman atau penguatan identitas sosial dalam komunitas digital (Wang et al., 2014).

Studi tersebut menemukan bahwa pengguna internet cenderung menyembunyikan identitas asli mereka ketika terlibat dalam perilaku tidak pantas secara daring. Namun, hubungan ini dimoderasi oleh faktor-faktor eksternal seperti kepastian hukuman (*punishment certainty*), keparahan hukuman (*punishment severity*), dan identitas sosial (*social identity*). Ketika pengguna menyadari bahwa perilaku mereka akan terdeteksi dan dihukum berat, atau ketika mereka memiliki identitas sosial yang kuat dalam komunitas daring, perilaku disinhibisi toksik cenderung berkurang (Wang et al., 2014).

Temuan-temuan ini menunjukkan bahwa anonimitas dan kontrol identitas beroperasi secara ambivalen: di satu sisi, anonimitas memfasilitasi kebebasan berekspresi dan pengungkapan diri yang positif; di sisi lain, dapat menurunkan hambatan psikologis terhadap perilaku tidak etis. Sejalan dengan studi ODE di Indonesia, anonimitas merupakan faktor teknis platform yang paling konsisten diidentifikasi sebagai pemicu utama ODE (Anggarani & Amalia, 2021; Bulan & Wulandari, 2021; Haqie et al., 2024; Lestari & Alwi, 2024; Samsiah & Utami Sumaryanti, 2023; Sudarja, 2025).

Lestari dan Alwi menemukan bahwa anonimitas dan berkurangnya akuntabilitas dalam interaksi online mendorong siswa MA pengguna Instagram untuk menunjukkan kecenderungan perilaku *cyberbullying* yang lebih tinggi (Lestari & Alwi, 2024). Temuan serupa dilaporkan oleh Bulan et al. yang menunjukkan bahwa penggunaan media sosial anonim memungkinkan remaja untuk menyembunyikan identitas asli

mereka, sehingga merasa lebih bebas untuk melakukan *cyberbullying* tanpa takut konsekuensi (Bulan & Wulandari, 2021).

Samsiah et al. menemukan bahwa anonimitas berkontribusi 50,4% terhadap varians perilaku *cyberbullying* pada remaja akhir di Bandung, menunjukkan efek yang sangat substansial (Samsiah & Utami Sumaryanti, 2023). Penelitian ini mengidentifikasi bahwa pelaku *cyberbullying* merasa aman dan terlindungi di balik layar, serta tidak perlu bertanggung jawab atas tindakannya ketika menggunakan akun anonim atau palsu.

Widyaning et al. menemukan bahwa anonimitas bersama dengan konsep diri menjelaskan 42,7% varians ODE pada remaja, menunjukkan bahwa meskipun anonimitas adalah faktor penting, interaksinya dengan karakteristik psikologis individu juga perlu dipertimbangkan (Haqie et al., 2024). Penelitian pada komunitas "marah-marah" di platform X menunjukkan bahwa anonimitas merupakan salah satu dimensi Suler yang berkontribusi pada tingkat ODE sedang (skor empiris 96,55) di kalangan anggota komunitas tersebut (Julianti et al., 2025).

Invisibilitas memperkuat dinamika tersebut melalui penghapusan *nonverbal feedback loops* yang dalam interaksi tatap muka berfungsi sebagai regulator utama perilaku sosial (Febriana & Fajrianthi, 2019; Siregar, 2026; S. Wu et al., 2017). Dalam absennya isyarat visual dan emosional, individu mengalami penurunan *empathetic attunement*, yang berimplikasi pada meningkatnya jarak psikologis terhadap lawan interaksi. Hal ini memfasilitasi proses dehumanisasi simbolik dan melemahkan sensitivitas terhadap dampak interpersonal dari komunikasi digital. Dalam kerangka ini, invisibilitas dapat dipahami sebagai mekanisme yang mengganggu *affective reciprocity* dalam interaksi sosial (Suler, 2004).

Febriana dan Fajrianthi menemukan bahwa invisibilitas, bersama dengan anonimitas dan asinkronitas, secara signifikan memengaruhi *cyber incivility* di lingkungan kerja Indonesia. Ketidakhadiran kontak mata, ekspresi wajah, dan bahasa tubuh mengurangi empati dan akuntabilitas sosial, memudahkan individu untuk terlibat dalam perilaku yang tidak sopan atau agresif (Febriana & Fajrianthi, 2019).

Siregar (2026). mengidentifikasi bahwa dimensi *invisibility* signifikan dalam konteks permainan daring kompetitif, di mana pemain tidak dapat melihat reaksi fisik lawan main mereka, sehingga lebih mudah untuk terlibat dalam *toxic behavior* seperti *verbal abuse* dan *misconduct*. Penelitian ini menunjukkan bahwa invisibilitas menciptakan jarak psikologis yang memfasilitasi dehumanisasi lawan interaksi.

Sementara itu, asinkronitas mengubah struktur temporal interaksi dengan memutus *real-time social feedback*, yang pada kondisi normal berfungsi sebagai mekanisme koreksi perilaku. Ketiadaan respons langsung memungkinkan individu untuk mengembangkan ekspresi kognitif dan emosional tanpa tekanan sosial yang bersifat immediat, sehingga memperbesar kemungkinan munculnya ekspresi ekstrem, baik dalam bentuk *benign* maupun *toxic disinhibition*. Dalam istilah Suler, kondisi ini memungkinkan terjadinya pola komunikasi yang menyerupai *emotional detachment* atau bahkan *emotional hit-and-run* (Suler, 2004).

Asinkronitas komunikasi, atau sifat tidak sinkron dari banyak interaksi online, memungkinkan individu untuk menghindari reaksi langsung dari lawan bicara dan merencanakan respons tanpa tekanan waktu (Febriana & Fajrianthi, 2019; Gumelar et al., 2024; Zahriyanto et al., 2025). Febriana dan Fajrianthi (2019) menemukan bahwa asinkronitas merupakan salah satu dari tiga faktor utama

(bersama anonimitas dan invisibilitas) yang memprediksi *cyber incivility* di tempat kerja.

Penelitian pada komunitas online menunjukkan bahwa asinkronitas memberikan kesempatan bagi individu untuk memposting konten provokatif atau agresif tanpa harus menghadapi konsekuensi sosial langsung (Julianti et al., 2025; Manalu, 2023). Manalu (2023) mengidentifikasi bahwa perpindahan dari forum terbuka ke forum tertutup dan penggunaan komunikasi asinkron memfasilitasi extreme speech dan ujaran kasar di kalangan remaja.

Selain faktor teknis tersebut, Suler (2004) menyebut ODE juga dipengaruhi oleh mekanisme psikologis yang lebih dalam, seperti *solipsistic introjection*, *dissociative imagination*, dan *Minimization of authority*. *Solipsistic introjection* merujuk pada kondisi di mana individu “menginternalisasi” lawan interaksi daring ke dalam ruang psikologisnya sendiri, sehingga komunikasi digital tidak lagi dirasakan sebagai interaksi dengan orang lain, melainkan sebagai dialog intrapsikis. Dalam komunikasi berbasis teks, ketiadaan isyarat visual dan vokal membuat individu secara tidak sadar “mengisi kekosongan” tersebut dengan imajinasi, misalnya dengan membayangkan suara, ekspresi, atau karakter lawan bicara.

Akibatnya, batas antara “diri” dan “orang lain” menjadi kabur, dan interaksi daring dipersepsikan seolah-olah terjadi dalam pikiran sendiri. Kondisi ini menurunkan hambatan sosial karena berbicara dengan “representasi mental” terasa lebih aman dibandingkan berhadapan langsung dengan individu nyata. Dalam konteks interaksi digital, mekanisme ini berkontribusi pada meningkatnya keterbukaan, tetapi juga dapat mendorong ekspresi impulsif dan tidak terfilter secara sosial (Suler, 2004).

Dissociative imagination mengacu pada kecenderungan individu untuk memandang ruang

digital sebagai domain yang terpisah dari realitas sehari-hari. Ruang digital seolah-olah merupakan “ruang imajinatif” atau bahkan “permainan sosial” yang tidak sepenuhnya terikat oleh norma dunia nyata. Dalam kondisi ini, individu memisahkan (*dissociate*) identitas dan perilaku daring dari identitas offline mereka. Akibatnya, tindakan yang dilakukan di ruang digital dianggap tidak memiliki konsekuensi nyata atau moral yang sama dengan dunia offline. Individu dapat merasa bahwa apa yang terjadi di dunia daring adalah bagian dari “peran” atau “karakter” yang dimainkan, sehingga tanggung jawab personal menjadi melemah (Suler, 2004).

Sedangkan *Minimization of authority* merujuk pada melemahnya pengaruh struktur otoritas dalam ruang digital akibat absennya penanda sosial seperti status, jabatan, simbol kekuasaan, maupun bahasa tubuh yang biasanya memperkuat hierarki dalam interaksi tatap muka. Dalam lingkungan daring, semua individu cenderung tampil dalam posisi yang relatif setara (*flattened hierarchy*), sehingga batas antara otoritas dan non-otoritas menjadi kabur. Kondisi ini membuat individu merasa lebih bebas untuk mengekspresikan opini, termasuk terhadap figur yang dalam konteks offline memiliki kekuasaan atau otoritas lebih tinggi. Di satu sisi, hal ini dapat mendorong partisipasi dan demokratisasi komunikasi; namun di sisi lain, juga dapat memfasilitasi perilaku tidak sopan, agresif, atau bahkan delegitimasi terhadap otoritas (Suler, 2004).

Di luar faktor struktural yang disebutkan oleh Suler, kajian terbaru menunjukkan bahwa ODE sangat dipengaruhi oleh konfigurasi psikologis individu, khususnya dalam bentuk *personality dispositions* dan mekanisme regulasi moral. Ciri kepribadian *Dark Triad* (Machiavellianism, psikopati, dan narsisisme) berfungsi sebagai *amplifying variables* yang

memperkuat kecenderungan individu untuk memanfaatkan kondisi disinhibisi dalam ruang digital. Individu dengan karakteristik ini cenderung menunjukkan rendahnya empati, tingginya orientasi instrumental, serta kecenderungan untuk mengeksploitasi celah anonimitas dalam interaksi daring (Rafly & Ningsih, 2025; B. Wu et al., 2022, 2023).

Lebih jauh, mekanisme *moral disengagement* berperan sebagai *cognitive mediator* yang menjembatani hubungan antara kondisi disinhibisi dan perilaku tidak etis. Dalam konteks ini, individu secara aktif merekonstruksi justifikasi moral atas tindakannya melalui proses rasionalisasi, distorsi konsekuensi, atau delegitimasi korban. Temuan empiris menunjukkan bahwa ODE secara konsisten meningkatkan kecenderungan *moral disengagement*, yang pada gilirannya memperkuat praktik seperti *cyber incivility*, agresi daring, dan bentuk-bentuk kekerasan digital lainnya (Gumelar et al., 2024).

Hal ini menandakan bahwa degradasi etika dalam ruang digital tidak hanya disebabkan oleh absennya kontrol eksternal, tetapi juga oleh transformasi dalam struktur kognisi moral individu. Intervensi yang berfokus pada penguatan standar moral dan peningkatan kesadaran akan konsekuensi perilaku daring dapat membantu mengurangi *moral disengagement* dan, pada gilirannya, mengurangi perilaku tidak etis (Gan et al., 2024; Gumelar et al., 2024; Sánchez-Hernández et al., 2024).

Namun demikian, efek ODE tidak bersifat deterministik. Literatur mengidentifikasi keberadaan *protective factors* seperti empati, regulasi emosi, dan *mindfulness* yang berfungsi sebagai *counter-regulatory mechanisms* terhadap disinhibisi (Soares et al., 2023; Syrjämäki et al., 2024). Individu dengan kapasitas empatik yang tinggi menunjukkan resistensi yang lebih besar

terhadap dorongan agresif, sementara kemampuan regulasi emosi memungkinkan individu mempertahankan kontrol diri dalam situasi komunikasi yang provokatif. Dalam hal ini, ODE dapat dipahami sebagai fenomena yang dimoderasi oleh kapasitas psikologis individu, bukan sebagai kondisi yang secara otomatis menghasilkan perilaku menyimpang.

Selain itu, faktor kontekstual seperti norma komunitas, afiliasi kelompok sebaya, serta motif penggunaan platform turut membentuk dinamika ODE. Zahriyanto et al., (2025) mengidentifikasi tantangan dan peluang Pancasila sebagai fondasi etika sosial di era digital. Penelitian ini menempatkan fenomena ODE dalam konteks nilai-nilai budaya Indonesia dan menunjukkan bahwa ada ketegangan antara nilai-nilai tradisional yang menekankan kesopanan, harmoni, dan gotong royong dengan perilaku online yang sering disinhibited dan individualistik.

Penelitian ini menunjukkan bahwa pendekatan untuk mengatasi ODE di Indonesia perlu mempertimbangkan konteks budaya lokal dan memanfaatkan nilai-nilai Pancasila sebagai kerangka etika untuk perilaku digital. Integrasi nilai-nilai lokal dalam pendidikan digital mungkin lebih efektif daripada pendekatan yang sepenuhnya mengadopsi model Barat.

Dengan demikian, *Online Disinhibition Effect* merupakan fenomena yang tidak dapat direduksi pada aspek psikologis individual semata, melainkan harus dipahami sebagai hasil dari interaksi antara *platform architecture*, dinamika psikologis, dan struktur sosial digital. Ruang digital dalam hal ini bertindak sebagai arena yang tidak netral, melainkan sebagai sistem yang secara aktif mengonstruksi ulang batas-batas etika, identitas, dan relasi sosial. Pemahaman terhadap kompleksitas ini menjadi krusial sebagai dasar untuk menganalisis

implikasi ODE terhadap etika kewargaan digital dalam masyarakat kontemporer.

Dimensi *Benign* dan *Toxic Disinhibition* dalam ODE

Berbagai studi tentang *Online Disinhibition Effect* (ODE) menunjukkan adanya perbandingan antara dimensi *benign disinhibition* (disinhibisi yang bersifat prososial atau positif) dan *toxic disinhibition* (disinhibisi yang bersifat merugikan).

Benign disinhibition merepresentasikan dimensi prososial dari *Online Disinhibition Effect*, yang ditandai dengan meningkatnya keterbukaan diri, ekspresi emosi yang autentik, serta munculnya perilaku empatik dalam interaksi digital. Suler menjelaskan bahwa dalam kondisi ini individu cenderung mengungkapkan pengalaman personal yang mendalam, termasuk emosi, ketakutan, dan keinginan yang dalam konteks tatap muka cenderung terhambat oleh norma sosial. Disinhibisi dalam bentuk ini tidak hanya bersifat ekspresif, tetapi juga reflektif, karena memungkinkan individu melakukan *working through* terhadap pengalaman intrapsikis serta mendorong proses *self-actualization* (Suler, 2004).

Sebaliknya, *toxic disinhibition* merepresentasikan manifestasi disinhibisi yang bersifat destruktif, yang ditandai dengan munculnya perilaku agresif, penggunaan bahasa kasar, ujaran kebencian, hingga ancaman yang dalam kondisi normal cenderung terhambat dalam interaksi tatap muka. Dimensi ini menjadi fokus dominan dalam literatur karena berkaitan langsung dengan berbagai fenomena problematik dalam ruang digital, seperti *cyberaggression*, *cyberbullying*, *trolling*, dan *hate speech* (Cheung et al., 2020).

Hasil penelusuran literatur pada berbagai basis data ilmiah menunjukkan bahwa terdapat

191 publikasi yang membahas ODE secara umum, dengan 120 publikasi mengkaji *benign disinhibition* dan 123 publikasi mengkaji *toxic disinhibition*. Meskipun secara kuantitatif jumlah publikasi pada kedua dimensi relatif seimbang, analisis terhadap orientasi penelitian menunjukkan adanya ketimpangan yang signifikan. Dimensi *toxic disinhibition* cenderung memperoleh perhatian lebih besar, terutama dalam konteks pengukuran, identifikasi risiko, serta pengembangan strategi mitigasi dan intervensi. Sebaliknya, *benign disinhibition* relatif kurang dieksplorasi secara empiris, meskipun memiliki implikasi penting dalam mendorong keterbukaan diri, dukungan sosial, serta penguatan etika dalam lingkungan digital.

Kecenderungan dominasi dimensi *Toxic* dalam penelitian dapat dijelaskan melalui dua faktor utama. Pertama, terdapat bias epistemik dalam praktik penelitian yang lebih memprioritaskan fenomena yang bersifat problematik, seperti agresi daring dan pelanggaran norma, karena lebih mudah dioperasionalisasikan dan diukur secara kuantitatif (Cheung et al., 2020). Perilaku seperti *cyberbullying* atau ujaran kebencian memiliki indikator yang lebih jelas dan dapat diamati secara langsung, sehingga lebih sering menjadi objek penelitian dibandingkan bentuk-bentuk perilaku prososial yang cenderung lebih subtil dan kontekstual.

Kedua, karakteristik *socio-technical affordances* dalam ruang digital secara struktural memang lebih sering berfungsi sebagai pemicu disinhibisi negatif. Faktor-faktor seperti anonimitas, invisibilitas, dan minimisasi otoritas terbukti menurunkan akuntabilitas, mengaburkan identitas sosial, serta melemahkan kontrol normatif dalam interaksi daring, sehingga lebih kondusif terhadap munculnya perilaku agresif dibandingkan perilaku prososial (Lapidot-Lefler

& Barak, 2012). Dalam konteks ini, dimensi *toxic disinhibition* tidak hanya merupakan hasil dari kecenderungan individual, tetapi juga refleksi dari desain dan logika operasional platform digital itu sendiri.

Meskipun demikian, dimensi *benign disinhibition* tetap menunjukkan potensi signifikan yang belum banyak dieksplorasi dalam literatur. Beberapa studi menunjukkan bahwa disinhibisi daring dapat memfasilitasi keterbukaan diri yang lebih mendalam, pembentukan relasi interpersonal yang lebih intim, serta dukungan sosial berbasis empati, terutama bagi individu yang mengalami hambatan dalam interaksi tatap muka (Martin, 2013; Suler, 2004). Dalam konteks tertentu, ODE bahkan dapat berfungsi sebagai mekanisme pemberdayaan, yang memungkinkan individu mengekspresikan identitas dan pengalaman personal secara lebih autentik dalam ruang digital.

Dengan demikian, sintesis temuan ini menunjukkan bahwa dominasi narasi mengenai dampak negatif ODE dalam literatur bukan semata-mata mencerminkan sifat inherent dari fenomena tersebut, melainkan merupakan hasil dari interaksi antara bias penelitian, perbedaan orientasi metodologis, serta karakteristik struktural ruang digital yang lebih kondusif terhadap ekspresi perilaku menyimpang. Kondisi ini berimplikasi langsung pada kecenderungan hasil penelitian yang lebih banyak menyoroti dampak negatif ODE, sebagaimana akan dibahas lebih lanjut pada bagian berikutnya mengenai implikasi disinhibisi daring terhadap etika digital.

Dampak Online Disinhibition Effect terhadap Perilaku dan Etika Digital

Berdasarkan sintesis terhadap temuan penelitian mengenai *Online Disinhibition Effect* (ODE) secara konsisten menunjukkan

kecenderungan yang lebih dominan pada dimensi *toxic disinhibition* dibandingkan *benign disinhibition*. Hal ini terlihat dari fokus empiris yang lebih banyak mengaitkan ODE dengan perilaku negatif seperti *trolling*, *cyberbullying*, agresi verbal, *toxic behavior*, hingga penyebaran hoaks dan pelanggaran norma komunikasi digital (Gan et al., 2024; Lai & Tsai, 2016; Soares et al., 2023; C. Wang et al., 2014; B. Wu et al., 2022, 2023). Dominasi temuan pada aspek toksik ini tidak semata-mata mencerminkan sifat inherent ODE yang cenderung negatif, melainkan juga dipengaruhi oleh bias dalam desain penelitian yang lebih menitikberatkan pada perilaku menyimpang yang lebih mudah diukur dan diamati, seperti agresi dan pelanggaran norma.

Selain itu, faktor struktural dalam ruang digital, seperti anonimitas, invisibilitas, dan asinkronitas, secara empiris terbukti lebih kuat dalam menurunkan akuntabilitas dan kontrol sosial, sehingga lebih sering memfasilitasi ekspresi perilaku agresif dibandingkan ekspresi positif (Lai & Tsai, 2016; Ramadhani & Merida, 2024). Di sisi lain, dimensi *benign disinhibition*, seperti peningkatan keterbukaan, dukungan emosional, atau ekspresi diri yang autentik, relatif kurang dieksplorasi dalam literatur, sehingga menghasilkan pemahaman yang cenderung parsial terhadap fenomena ODE. Dengan demikian, kecenderungan temuan yang mengarah pada *toxic disinhibition* dapat dipahami sebagai hasil dari interaksi antara karakteristik struktural ruang digital yang memang memfasilitasi pelemahan kontrol sosial, serta keterbatasan epistemik dalam penelitian yang lebih berfokus pada manifestasi negatif dibandingkan eksplorasi potensi positif dari disinhibisi daring.

Implikasi dari kecenderungan tersebut terlihat jelas pada pola temuan empiris yang berkembang dalam literatur, di mana dampak *Online Disinhibition Effect* (ODE) lebih banyak

dikaitkan dengan berbagai bentuk degradasi etika dalam interaksi digital. *Cyberbullying* merupakan dampak etika yang paling konsisten dan sering dilaporkan dari ODE dalam penelitian Indonesia (Anggarani & Amalia, 2021; Bulan & Wulandari, 2021; Juliati et al., 2025; Lestari & Alwi, 2024; Samsiah & Utami Sumaryanti, 2023; Shabira & Lestari, 2024). Lestari & Alwi, (2024) menemukan hubungan signifikan antara ODE dengan kecenderungan perilaku cyberbullying pada siswa MA pengguna Instagram, dengan manifestasi berupa perilaku mengancam dan agresi verbal.

Sintesis terhadap berbagai studi menunjukkan bahwa *Online Disinhibition Effect* (ODE) berperan sebagai prediktor kuat dalam perilaku *cyberbullying*, terutama melalui mekanisme penurunan kontrol diri dan pelemahan hambatan psikologis terhadap agresi (Lai & Tsai, 2016). Dalam kondisi disinhibisi, individu cenderung mengalami distorsi persepsi terhadap konsekuensi tindakan, sehingga lebih mudah mengekspresikan perilaku menyerang tanpa mempertimbangkan dampak terhadap korban.

Lebih jauh, hubungan antara ODE dan *cyberbullying* tidak bersifat langsung, melainkan dimediasi oleh faktor kognitif dan afektif, terutama *moral disengagement* dan empati. Individu yang mengalami disinhibisi cenderung mengaktifkan mekanisme pembenaran moral, seperti merasionalisasi agresi atau mereduksi persepsi terhadap penderitaan korban, sehingga perilaku *cyberbullying* dapat dilakukan tanpa konflik moral yang signifikan (Martin, 2013). Sebaliknya, empati berfungsi sebagai faktor protektif yang mampu melemahkan pengaruh ODE terhadap agresi daring, di mana individu dengan tingkat empati yang tinggi menunjukkan kecenderungan lebih rendah untuk terlibat dalam *cyberbullying*, bahkan dalam kondisi disinhibisi.

Selain faktor individual, dinamika sosial juga memperkuat efek ODE dalam konteks *cyberbullying*. Afiliasi dengan teman sebaya yang menyimpang terbukti memiliki hubungan positif dengan perilaku agresif daring, dan ODE berperan sebagai moderator yang memperkuat pengaruh tersebut (Yang et al., 2021). Dalam kondisi disinhibisi, norma kelompok yang menyimpang menjadi lebih mudah diinternalisasi, sehingga individu cenderung menyesuaikan perilakunya dengan lingkungan sosial digital yang permisif terhadap agresi. Hal ini menunjukkan bahwa ODE tidak hanya memfasilitasi perilaku individual, tetapi juga memperkuat transmisi norma negatif dalam komunitas daring.

Di sisi lain, karakteristik platform digital turut memperbesar risiko *cyberbullying*. Fitur seperti anonimitas dan rendahnya visibilitas identitas memperlemah akuntabilitas serta mengurangi sensitivitas terhadap dampak interpersonal, sehingga menciptakan kondisi yang kondusif bagi munculnya agresi daring (Juliati et al., 2025; Lai & Tsai, 2016). Dalam konteks ini, desain platform tidak hanya menjadi medium interaksi, tetapi juga berfungsi sebagai faktor struktural yang memediasi dan memperkuat efek disinhibisi.

Dengan demikian, *cyberbullying* dalam kerangka ODE dapat dipahami sebagai hasil dari interaksi antara mekanisme psikologis (disinhibisi dan *moral disengagement*), faktor afektif (empati), dinamika sosial (peer influence), serta karakteristik struktural platform digital. Kompleksitas ini menunjukkan bahwa *cyberbullying* bukan sekadar perilaku individual yang menyimpang, melainkan fenomena multidimensional yang berakar pada kondisi interaksi digital itu sendiri (Juliati et al., 2025; Lai & Tsai, 2016; Yang et al., 2021).

Sementara itu, *Trolling* juga menjadi salah satu manifestasi paling umum dari *toxic disinhibition*, yang ditandai oleh tindakan sengaja untuk memprovokasi, mengganggu, atau menyinggung orang lain dalam interaksi digital. Sintesis literatur menunjukkan bahwa *Online Disinhibition Effect* (ODE) berperan sebagai pendorong utama perilaku ini melalui pelemahan kontrol diri dan penurunan sensitivitas terhadap norma sosial, sehingga individu lebih mudah mengekspresikan agresi simbolik tanpa hambatan interpersonal yang biasanya hadir dalam interaksi tatap muka (B. Wu et al., 2022, 2023).

Hubungan antara ODE dan *trolling* diperkuat oleh faktor kepribadian, khususnya ciri *Dark Triad*. Individu dengan tingkat Machiavellianism dan psikopati yang tinggi menunjukkan kecenderungan lebih besar untuk memanfaatkan kondisi disinhibisi sebagai sarana manipulasi, provokasi, dan eksploitasi interaksi sosial daring (B. Wu et al., 2022, 2023). Dalam konteks ini, ODE tidak hanya membuka peluang bagi perilaku *trolling*, tetapi juga berfungsi sebagai *amplifier* bagi predisposisi kepribadian yang bersifat antisosial.

Selain faktor kepribadian, dimensi motivasional juga memainkan peran penting dalam menjelaskan munculnya *trolling*. Studi menunjukkan bahwa perilaku ini sering kali didorong oleh motif rekreasi dan pencarian penghargaan sosial, di mana individu memperoleh kepuasan dari reaksi emosional yang ditimbulkan pada orang lain (Soares et al., 2023). Dalam ekosistem digital yang berbasis *engagement*, respons negatif sekalipun dapat berfungsi sebagai bentuk penguatan (*reinforcement*), sehingga perilaku *trolling* tidak hanya dipicu oleh disinhibisi, tetapi juga dipertahankan oleh dinamika interaksi platform itu sendiri. Sebaliknya, empati kognitif berfungsi sebagai faktor protektif yang mampu menekan

kecenderungan tersebut, sehingga individu dengan empati rendah menjadi lebih rentan terhadap perilaku *trolling* (Soares et al., 2023).

Dampak dari *trolling* tidak terbatas pada interaksi individual, tetapi juga berimplikasi pada kualitas ekosistem komunikasi digital secara keseluruhan. Kehadiran perilaku provokatif yang berulang dapat menurunkan rasa aman pengguna, menghambat partisipasi, serta merusak kualitas diskursus dalam ruang publik digital (Soares et al., 2023). Dalam jangka panjang, normalisasi *trolling* berpotensi menciptakan budaya komunikasi yang permisif terhadap agresi simbolik, yang pada akhirnya menggeser standar etika interaksi daring.

Dengan demikian, *trolling* dalam kerangka ODE dapat dipahami sebagai hasil interaksi antara disinhibisi psikologis, predisposisi kepribadian, motif sosial, serta struktur insentif dalam platform digital. Kompleksitas ini menunjukkan bahwa *trolling* bukan sekadar perilaku individu yang menyimpang, melainkan fenomena yang diproduksi dan direproduksi dalam dinamika interaksi digital itu sendiri (Soares et al., 2023; B. Wu et al., 2022).

Lebih lanjut, penyebaran hoaks dan disinformasi merupakan salah satu dampak paling luas dari *Online Disinhibition Effect* (ODE) karena berimplikasi langsung pada kualitas ruang publik digital dan praktik kewargaan. Sintesis temuan menunjukkan bahwa *toxic disinhibition* berkontribusi signifikan terhadap kecenderungan individu untuk membagikan informasi tanpa verifikasi, terutama melalui mekanisme penurunan kontrol diri dan melemahnya akuntabilitas dalam kondisi anonimitas (Matondang & Afdal, 2023). Dalam konteks ini, individu tidak hanya bertindak lebih impulsif, tetapi juga mengalami penurunan sensitivitas terhadap konsekuensi sosial dari tindakan berbagi informasi.

Secara mekanistik, ODE bekerja melalui dua jalur utama. Pertama, jalur kognitif-afektif, di mana disinhibisi menurunkan *epistemic vigilance*, yakni kehati-hatian dalam mengevaluasi kebenaran informasi, sehingga individu lebih mudah mempercayai dan menyebarkan konten yang selaras dengan emosi atau keyakinan pribadi. Kedua, jalur normatif, di mana anonimitas dan invisibilitas mengaburkan tanggung jawab sosial, sehingga tindakan berbagi informasi tidak lagi dipersepsikan sebagai tindakan bermuatan etis yang berdampak pada orang lain (Matondang & Afdal, 2023). Kombinasi kedua jalur ini menjelaskan mengapa perilaku berbagi hoaks sering terjadi tanpa niat jahat, melainkan sebagai hasil dari impulsivitas yang dimediasi oleh kondisi psikologis dalam interaksi digital.

Dalam konteks Indonesia, dinamika ini menjadi semakin signifikan karena penyebaran hoaks telah berdampak pada kohesi sosial, proses politik, dan kesehatan publik, terutama pada masa krisis seperti pandemi COVID-19 (Matondang & Afdal, 2023). ODE mempercepat sirkulasi disinformasi dengan menurunkan ambang refleksi kritis, sehingga ruang publik digital cenderung dipenuhi oleh arus informasi yang tidak terverifikasi dan berpotensi menyesatkan.

Berbeda dengan dominasi temuan pada *toxic disinhibition*, dimensi *benign disinhibition* menunjukkan bahwa ODE juga memiliki potensi konstruktif dalam interaksi digital, meskipun eksplorasinya dalam literatur relatif terbatas. Dampak positif ini terutama termanifestasi dalam peningkatan keterbukaan diri, dukungan sosial, serta pemanfaatannya dalam konteks pendidikan dan terapeutik.

Salah satu dampak paling konsisten dari *benign disinhibition* adalah meningkatnya *self-disclosure* dan keterbukaan emosional dalam komunikasi daring. Lingkungan digital yang

ditandai oleh anonimitas dan invisibilitas menurunkan hambatan sosial yang biasanya membatasi ekspresi diri dalam interaksi tatap muka, sehingga individu lebih leluasa mengungkapkan pengalaman personal, emosi, dan kerentanan psikologis (Clark-Gordon et al., 2019; Lapidot-Lefler & Barak, 2012). Temuan eksperimental menunjukkan bahwa kombinasi anonimitas dan invisibilitas secara signifikan meningkatkan tingkat pengungkapan diri, yang mengindikasikan bahwa *socio-technical affordances* tertentu dapat secara langsung memfasilitasi ekspresi emosional yang lebih autentik (Lapidot-Lefler & Barak, 2012).

Dari perspektif kognitif-afektif, kondisi ini dapat dipahami sebagai penurunan *social evaluation anxiety*, yang memungkinkan individu mengekspresikan diri tanpa takut terhadap penilaian sosial. Dalam konteks tertentu, peningkatan *self-disclosure* ini berpotensi mendukung pemrosesan emosi dan refleksi diri, sehingga berkontribusi pada kesejahteraan psikologis individu (Clark-Gordon et al., 2019).

Selain keterbukaan emosional, *benign disinhibition* juga berpotensi memfasilitasi dukungan sosial dan perilaku prososial dalam ruang digital. Suler menyatakan bahwa disinhibisi dapat mendorong individu untuk menunjukkan empati, kepedulian, dan bahkan tindakan altruistik yang dalam kondisi offline mungkin terhambat oleh norma sosial atau kecanggungan interpersonal (Suler, 2004). Dalam konteks ini, ODE dapat berfungsi sebagai mekanisme yang memperluas kapasitas individu untuk membangun relasi yang suportif dan bermakna.

Namun demikian, temuan empiris menunjukkan bahwa hubungan antara *benign disinhibition* dan perilaku prososial tidak selalu konsisten. Studi eksperimental menunjukkan bahwa peningkatan *self-disclosure* tidak secara

otomatis diikuti oleh peningkatan perilaku prososial, yang mengindikasikan bahwa ekspresi diri dan tindakan prososial merupakan dua proses yang berbeda secara psikologis (Lapidot-Lefler & Barak, 2012). Hal ini menunjukkan bahwa potensi prososial dari ODE bersifat kondisional dan bergantung pada faktor lain seperti empati, norma kelompok, dan konteks interaksi.

Dimensi *benign disinhibition* juga memiliki implikasi signifikan dalam konteks pendidikan dan terapeutik. Dalam lingkungan pembelajaran, disinhibisi daring memungkinkan siswa untuk berpartisipasi lebih aktif, terutama bagi individu yang cenderung pemalu atau mengalami hambatan dalam komunikasi tatap muka. Interaksi digital membuka ruang bagi ekspresi pendapat yang lebih bebas serta memperluas jaringan sosial di luar kelompok pertemanan yang sudah ada (Martin, 2013).

Dalam konteks terapeutik, *benign disinhibition* berperan dalam mempercepat proses *self-disclosure*, yang merupakan elemen kunci dalam hubungan terapeutik. Individu lebih mudah mengungkapkan pengalaman emosional yang sensitif dalam lingkungan daring, sehingga dapat mendukung proses pemulihan psikologis dan pencarian dukungan emosional (Clark-Gordon et al., 2019). Dalam hal ini, ODE berfungsi sebagai fenomena perilaku sekaligus juga sebagai mekanisme yang dapat dimanfaatkan secara strategis dalam praktik intervensi psikologis.

Meskipun memiliki potensi konstruktif, temuan menunjukkan bahwa dampak *benign disinhibition* masih kurang tereksplorasi dibandingkan dimensi toksik. Hal ini disebabkan oleh bias penelitian serta dampak positif ODE cenderung lebih subtil, kontekstual, dan sulit dioperasionalisasikan dalam desain penelitian kuantitatif. Akibatnya, kontribusi *benign disinhibition* terhadap pembentukan interaksi digital yang sehat seringkali kurang terlihat dalam

literatur, meskipun memiliki peran penting dalam mendukung kesejahteraan psikologis dan kualitas relasi sosial. Dengan demikian, ODE memiliki dua dimensi yang menghasilkan degradasi etika digital sekaligus memiliki potensi sebagai mekanisme yang dapat memperkuat dimensi prososial dalam interaksi daring. Realisasi potensi tersebut sangat bergantung pada kondisi psikologis individu, norma sosial yang berlaku, serta desain lingkungan digital yang digunakan.

Keterbatasan Penelitian dalam Menggali Fenomena Online Disinhibition Effect

Sintesis terhadap berbagai studi menunjukkan bahwa penelitian mengenai *Online Disinhibition Effect* (ODE) masih menghadapi sejumlah keterbatasan yang mencakup aspek sampel, metodologi, konteks, serta pengukuran. Keterbatasan tersebut tercermin dalam kecenderungan penggunaan sampel yang homogen, terutama mahasiswa dan remaja, yang membatasi generalisasi temuan ke populasi yang lebih beragam, dominasi desain *cross-sectional* dan data *self-report* yang membatasi inferensi kausal, serta kurangnya eksplorasi terhadap variasi konteks budaya dan karakteristik platform digital yang berbeda. Selain itu, variasi dalam instrumen pengukuran ODE dan konstruk terkait juga menyulitkan konsistensi dan komparabilitas hasil lintas studi. Secara keseluruhan, kondisi ini menunjukkan bahwa pemahaman yang ada mengenai ODE masih berkembang dalam batasan metodologis dan kontekstual tertentu, yang perlu diperhatikan dalam menafsirkan temuan penelitian.

Keterbatasan metodologis yang paling konsisten dalam penelitian ODE di Indonesia adalah dominasi desain *cross-sectional* atau potong-lintang (Bulan & Wulandari, 2021; Haqie et al., 2024; Julianti et al., 2025; Samsiah & Utami Sumaryanti, 2023) dan data *self-report*,

yang memiliki keterbatasan metodologis signifikan. Desain ini mengukur variabel pada satu titik waktu, yang membatasi kemampuan untuk membuat inferensi kausal yang kuat tentang hubungan antara faktor-faktor ODE dan dampak etika (Bulan & Wulandari, 2021; Haqie et al., 2024). Meskipun banyak studi menggunakan analisis mediasi dan moderasi untuk menguji mekanisme yang mendasari, hubungan kausal yang sebenarnya hanya dapat ditetapkan melalui desain longitudinal atau eksperimental.

Penggunaan data *self-report* dalam penelitian ODE juga rentan terhadap berbagai bias, seperti bias desirabilitas sosial, bias ingatan, dan bias respons, yang berpotensi mengurangi akurasi pelaporan perilaku. Dalam konteks ini, partisipan cenderung mereduksi atau tidak melaporkan secara penuh perilaku tidak etis yang mereka lakukan, baik karena kekhawatiran terhadap penilaian sosial maupun keterbatasan kesadaran reflektif terhadap perilaku daring mereka. Kondisi ini menyebabkan kemungkinan terjadinya *underreporting* terhadap praktik seperti *cyberbullying*, *trolling*, dan pelanggaran norma komunikasi digital.

Upaya untuk mengatasi keterbatasan tersebut telah dilakukan melalui penggunaan desain campuran dan pendekatan eksperimental, seperti yang ditunjukkan dalam studi Wang dan rekan yang menggabungkan survei dan eksperimen untuk menguji hubungan antarvariabel (C. Wang et al., 2014). Namun demikian, mayoritas penelitian masih bergantung pada desain *cross-sectional* berbasis *self-report*, sehingga keterbatasan ini tetap menjadi karakter dominan dalam literatur ODE.

Keterbatasan metodologis ini menunjukkan pentingnya pengembangan desain penelitian yang lebih robust, khususnya melalui pendekatan longitudinal untuk menangkap dinamika

perubahan perilaku dari waktu ke waktu, serta desain eksperimental untuk menguji hubungan kausal secara lebih meyakinkan. Selain itu, integrasi metode alternatif seperti observasi perilaku daring dan analisis *digital trace data* menjadi penting untuk meningkatkan validitas pengukuran dan mengurangi bias *self-report* (Gan et al., 2024; Soares et al., 2023; C. Wang et al., 2014).

Selanjutnya, variasi dalam skala dan instrumen pengukuran menjadi keterbatasan signifikan lain dalam penelitian ODE, yang berdampak pada rendahnya konsistensi dan komparabilitas temuan lintas studi. Berbagai penelitian menggunakan pendekatan operasionalisasi yang berbeda, seperti *Online Disinhibition Scale* yang dikembangkan oleh Udris maupun instrumen yang dikembangkan oleh Cheung dan rekan (Cheung et al., 2020). Perbedaan ini di satu sisi mencerminkan keragaman pendekatan konseptual, tetapi di sisi lain berimplikasi langsung pada perbedaan hasil penelitian, sehingga menyulitkan upaya sintesis dan meta-analisis secara sistematis.

Ketidakkonsistenan pengukuran juga terjadi pada konstruk-konstruk terkait, seperti *moral disengagement*, *Dark Triad*, dan empati. Sebagian studi menggunakan instrumen yang telah tervalidasi secara luas, sementara studi lainnya mengembangkan alat ukur sendiri sesuai kebutuhan penelitian. Variasi ini berpotensi memengaruhi reliabilitas dan validitas temuan, serta menimbulkan kesulitan dalam membandingkan hasil antar studi yang menggunakan indikator dan skala yang tidak seragam.

Upaya untuk mengatasi fragmentasi ini telah dilakukan melalui pengembangan kerangka pengukuran yang lebih komprehensif, seperti yang diusulkan oleh Cheung dan rekan, yang memasukkan dimensi-dimensi utama ODE

seperti anonimitas, invisibilitas, dan asinkronisitas dalam satu model pengukuran terpadu (Cheung et al., 2020). Namun demikian, adopsi terhadap pendekatan ini dalam penelitian selanjutnya masih terbatas, sehingga praktik pengukuran ODE tetap bersifat terfragmentasi.

Kondisi ini menunjukkan perlunya standarisasi dalam pengukuran ODE dan konstruk terkait, baik melalui pengembangan instrumen yang terintegrasi maupun penggunaan skala yang telah tervalidasi secara konsisten. Selain itu, transparansi dalam pelaporan properti psikometrik instrumen, seperti reliabilitas dan validitas, menjadi penting untuk memastikan kualitas pengukuran serta memfasilitasi evaluasi dan replikasi penelitian di masa mendatang (Cheung et al., 2020; Gan et al., 2024; B. Wu et al., 2022).

Keterbatasan lain dalam penelitian ODE berkaitan dengan aspek kontekstual dan budaya, yang memengaruhi tingkat generalisasi temuan. Studi-studi yang dianalisis berasal dari berbagai konteks geografis, seperti Tiongkok, Amerika Serikat, Kanada, Indonesia, dan negara-negara Eropa, yang masing-masing memiliki norma sosial, nilai budaya, dan praktik komunikasi yang berbeda (Gan et al., 2024; Gumelar et al., 2024; Soares et al., 2023; B. Wu et al., 2022). Variasi ini berpotensi memengaruhi bagaimana individu mengalami, menafsirkan, dan merespons disinhibisi dalam interaksi digital.

Perbedaan orientasi budaya, misalnya antara budaya kolektivistis dan individualis, menunjukkan implikasi yang signifikan terhadap manifestasi ODE. Budaya kolektivistis cenderung menekankan harmoni sosial dan konformitas, yang berpotensi memoderasi ekspresi perilaku tidak etis, sementara budaya individualis lebih mendorong ekspresi diri dan otonomi, yang dapat memperbesar peluang munculnya disinhibisi dalam bentuk yang lebih eksplisit. Namun

demikian, sebagian besar penelitian belum secara sistematis mengintegrasikan variabel budaya dalam desain analisis, sehingga pemahaman mengenai peran budaya dalam membentuk dinamika ODE masih terbatas.

Selain faktor budaya, karakteristik platform digital juga menjadi variabel kontekstual yang penting. Perbedaan fitur platform, seperti tingkat anonimitas, visibilitas identitas, serta norma komunitas, dapat memengaruhi tingkat dan bentuk disinhibisi yang muncul. Platform yang memungkinkan anonimitas tinggi cenderung lebih kondusif terhadap ekspresi toxic disinhibition, dibandingkan dengan platform yang menuntut identitas nyata dan akuntabilitas sosial. Namun, literatur yang ada masih jarang melakukan analisis komparatif lintas platform secara sistematis, sehingga hubungan antara desain platform dan variasi ODE belum sepenuhnya terpetakan.

Dengan demikian, keterbatasan kontekstual ini menunjukkan bahwa ODE merupakan fenomena yang bersifat situasional dan tidak dapat dilepaskan dari lingkungan sosial dan teknologi tempat ia berlangsung. Oleh karena itu, penelitian di masa depan perlu mengembangkan pendekatan komparatif lintas budaya dan lintas platform untuk memperoleh pemahaman yang lebih kontekstual dan komprehensif terhadap dinamika disinhibisi dalam interaksi digital (Gan et al., 2024; Gumelar et al., 2024; Juliati et al., 2025; Soares et al., 2023; B. Wu et al., 2022).

Salah satu keterbatasan yang paling menonjol dalam penelitian Online ODE adalah dominasi penggunaan sampel mahasiswa, yang berdampak pada terbatasnya generalisasi temuan. Sebagian besar studi menggunakan mahasiswa sebagai partisipan (Gan et al., 2024; C. Wang et al., 2014; B. Wu et al., 2022), sehingga representasi fenomena ODE cenderung didasarkan pada kelompok dengan karakteristik

demografis yang relatif homogen, seperti tingkat pendidikan yang lebih tinggi, akses teknologi yang memadai, serta konteks sosial yang spesifik dalam lingkungan akademik.

Karakteristik tersebut berpotensi memengaruhi cara individu mengalami dan merespons disinhibisi daring. Mahasiswa, misalnya, mungkin memiliki tingkat literasi digital dan kesadaran normatif yang lebih tinggi, yang dapat memoderasi ekspresi perilaku tidak etis dalam ruang digital. Sebaliknya, kelompok non-mahasiswa, seperti remaja sekolah atau dewasa muda yang telah bekerja, memiliki kondisi sosial, pengalaman hidup, dan dinamika interaksi yang berbeda, yang dapat menghasilkan pola ODE yang tidak sepenuhnya sejalan dengan temuan pada populasi mahasiswa.

Keterbatasan ini menunjukkan bahwa pemahaman mengenai ODE masih bersifat parsial dan berpotensi bias terhadap konteks kelompok tertentu. Oleh karena itu, penelitian di masa depan perlu mengembangkan desain sampling yang lebih representatif dan inklusif, dengan melibatkan berbagai kelompok usia, latar belakang sosial-ekonomi, serta konteks sosial yang beragam, guna memperoleh gambaran yang lebih komprehensif mengenai dinamika disinhibisi dalam interaksi digital (Gan et al., 2024; Soares et al., 2023; C. Wang et al., 2014; B. Wu et al., 2022).

Secara keseluruhan, sintesis literatur menunjukkan bahwa penelitian mengenai ODE masih dihadapkan pada sejumlah keterbatasan yang bersifat metodologis, kontekstual, dan konseptual, yang mencakup dominasi sampel yang tidak representatif, ketergantungan pada desain cross-sectional dan data self-report, terbatasnya eksplorasi lintas budaya dan platform digital, serta variasi instrumen pengukuran yang menghambat komparabilitas temuan. Keterbatasan-keterbatasan ini selain

memengaruhi validitas dan generalisasi hasil penelitian, juga berkontribusi terhadap terbentuknya pemahaman yang cenderung parsial terhadap fenomena ODE, khususnya dalam menyoroti dimensi *toxic disinhibition* dibandingkan aspek prososialnya. Oleh karena itu, diperlukan pengembangan penelitian yang lebih beragam dan integratif, baik melalui pendekatan longitudinal dan eksperimental, penggunaan metode pengukuran alternatif, maupun eksplorasi konteks sosial-budaya yang lebih luas, guna menghasilkan pemahaman yang lebih komprehensif dan seimbang sebagai dasar dalam merumuskan implikasi terhadap etika kewargaan digital.

Implikasi ODE dalam Perspektif Etika Kewargaan Digital

Temuan mengenai ODE pada penelitian ini menunjukkan bahwa fenomena ini memiliki implikasi yang lebih luas terhadap pembentukan etika kewargaan digital. Salah satu implikasi utama adalah perlunya integrasi pemahaman mengenai ODE dalam pendidikan kewarganegaraan untuk membangun etika digital, khususnya untuk membantu remaja menyadari bahwa karakteristik algoritma ruang digital seperti anonimitas, invisibilitas, dan asinkronitas dapat mendistorsi persepsi tanggung jawab dan memfasilitasi perilaku yang menyimpang dari norma etika di kehidupan offline (Juliati et al., 2025; Matondang & Afdal, 2023; Sudarja, 2025). Kesadaran ini penting untuk membangun prinsip akuntabilitas digital, di mana individu tetap memandang tindakan daring sebagai bagian dari praktik kewargaan yang memiliki konsekuensi sosial.

Selain itu, temuan mengenai peran kontrol diri sebagai faktor protektif terhadap *toxic disinhibition* menunjukkan bahwa pendidikan etika digital perlu diarahkan pada penguatan

kapasitas regulasi diri dan pengelolaan emosi. Kemampuan untuk menahan impuls, mempertimbangkan dampak jangka panjang, serta mengelola respons emosional dalam interaksi daring merupakan bagian dari kompetensi pendidikan kewarganegaraan yang esensial (Bulan & Wulandari, 2021; Haqie et al., 2024; Khairani & Guspa, 2025). Dengan demikian, pendidikan kewarganegaraan perlu melampaui pendekatan kognitif-informatif dan mengintegrasikan dimensi afektif serta psikologis dalam membentuk karakter digital yang bertanggungjawab.

Lebih jauh, mekanisme *moral disengagement* yang ditemukan dalam berbagai studi menunjukkan bahwa individu dapat merasionalisasi perilaku tidak etis dalam ruang digital tanpa mengalami konflik moral yang signifikan. Oleh karena itu, pendidikan etika kewarganegaraan digital perlu secara eksplisit menargetkan proses kognitif ini dengan membantu individu mengenali bentuk-bentuk rasionalisasi moral serta mengembangkan konsistensi nilai antara interaksi online dan offline (Gumelar et al., 2024). Pendekatan ini penting untuk mencegah fragmentasi moral, di mana individu menerapkan standar etika yang berbeda dalam konteks digital.

Implikasi lainnya adalah pentingnya pendekatan ekosistem dalam pendidikan kewarganegaraan, yang tidak hanya berfokus pada interaksi individu di sekolah, tetapi juga melibatkan keluarga sebagai agen sosialisasi utama. Temuan yang menunjukkan pengaruh pola asuh terhadap perilaku online remaja menegaskan bahwa orang tua memiliki peran strategis dalam membentuk kebiasaan digital yang sehat dan etis (Khairani & Guspa, 2025). Oleh karena itu, program literasi digital perlu mengintegrasikan edukasi bagi orang tua, sehingga mereka mampu memahami dinamika

interaksi digital dan memberikan pendampingan yang adaptif terhadap perkembangan teknologi.

Di Indonesia, penguatan etika kewarganegaraan digital juga perlu mempertimbangkan dimensi kultural dengan mengintegrasikan nilai-nilai lokal sebagai kerangka normatif. Nilai-nilai seperti gotong royong, kesopanan, dan harmoni sosial yang terkandung dalam Pancasila dapat menjadi landasan dalam membangun praktik komunikasi digital yang etis dan inklusif (Zahriyanto et al., 2025). Pendekatan yang kontekstual ini penting untuk memastikan bahwa pendidikan kewarganegaraan tidak bersifat normatif-abstrak, tetapi relevan dengan realitas sosial dan budaya pengguna.

Dengan demikian, implikasi ODE terhadap etika kewarganegaraan digital menunjukkan bahwa tantangan utama adalah membangun kapasitas individu untuk menjadi warga digital yang reflektif, bertanggung jawab, dan beretika dalam ruang publik digital yang semakin kompleks.

Selain implikasi pada pendidikan etika kewarganegaraan, temuan mengenai ODE juga menunjukkan pentingnya peran platform digital sebagai aktor struktural dalam membentuk perilaku pengguna. Karakteristik *socio-technical affordances* seperti anonimitas, invisibilitas, dan asinkronitas turut memediasi munculnya *toxic disinhibition*. Oleh karena itu, platform media sosial dan game online memiliki tanggung jawab strategis dalam merancang lingkungan digital yang dapat menjaga kualitas etika komunikasi (Juliati et al., 2025; Siregar, 2026).

Salah satu aspek kunci adalah pengelolaan anonimitas. Meskipun anonimitas memiliki fungsi penting dalam melindungi privasi dan kebebasan berekspresi, temuan literatur menunjukkan bahwa kondisi anonim juga dapat menurunkan akuntabilitas dan meningkatkan risiko perilaku tidak etis. Oleh karena itu, diperlukan pendekatan yang lebih seimbang,

seperti penerapan sistem verifikasi identitas opsional, diferensiasi tingkat anonimitas berdasarkan konteks, atau mekanisme reputasi digital yang memungkinkan akuntabilitas tanpa sepenuhnya menghilangkan privasi pengguna (Anggarani & Amalia, 2021; Bulan & Wulandari, 2021; Haqie et al., 2024; Juliati et al., 2025; Lestari & Alwi, 2024; Sudarja, 2025).

Selain itu, desain platform perlu diarahkan untuk mengurangi invisibilitas sosial yang menjadi pemicu utama disinhibisi. Hal ini dapat dilakukan dengan meningkatkan visibilitas konsekuensi sosial dari tindakan pengguna, seperti umpan balik terhadap dampak perilaku, serta fitur yang mendorong empati dengan mengingatkan bahwa interaksi digital melibatkan individu nyata [5], [9], [18]. Pendekatan ini penting untuk mengembalikan dimensi sosial dalam komunikasi digital yang cenderung tereduksi dalam interaksi berbasis online.

Dari sisi tata kelola, penguatan sistem moderasi menjadi elemen krusial dalam mengendalikan perilaku toksik. Kombinasi antara moderasi otomatis berbasis kecerdasan buatan dan moderasi manusia diperlukan untuk memastikan deteksi yang efisien sekaligus sensitif terhadap konteks (Febriana & Fajrianti, 2019; Siregar, 2026). Di samping itu, sistem pelaporan yang responsif dan transparan dapat meningkatkan kepercayaan pengguna serta memperkuat norma komunitas yang sehat.

Dalam konteks game online, desain berbasis insentif juga memainkan peran penting dalam membentuk perilaku pengguna. Penerapan sistem penghargaan untuk perilaku positif, serta konsekuensi yang konsisten terhadap pelanggaran, dapat membantu menggeser norma interaksi ke arah yang lebih konstruktif (Julianti et al., 2025).

Terakhir, platform juga perlu menyediakan alat bagi pengguna untuk mengelola pengalaman

digital mereka secara mandiri, seperti fitur kontrol privasi, pemblokiran, serta edukasi terkait keamanan dan kesejahteraan digital (Julianti et al., 2025; Lestari & Alwi, 2024; Siregar, 2026). Hal ini menunjukkan bahwa mitigasi ODE tidak hanya bergantung pada perubahan perilaku individu, tetapi juga pada desain dan regulasi ekosistem digital secara keseluruhan.

Kesimpulan

Berdasarkan sintesis terhadap 35 artikel yang memenuhi kriteria inklusi, penelitian ini menunjukkan bahwa *Online Disinhibition Effect* (ODE) merupakan fenomena yang terbentuk melalui interaksi antara karakteristik lingkungan digital, mekanisme psikologis individu, dan kondisi sosial-kontekstual. Karakteristik ruang digital berupa anonimitas, invisibilitas, dan asinkronitas menjadi faktor yang paling konsisten dikaitkan dengan melemahnya hambatan sosial dan akuntabilitas dalam interaksi daring, sementara *moral disengagement*, kontrol diri, empati, karakter kepribadian, serta pengaruh kelompok turut menentukan arah dan intensitas manifestasi ODE. Dengan demikian, ODE dalam literatur tidak menunjukkan pola sebagai respons psikologis yang berdiri sendiri, melainkan sebagai proses yang berlangsung dalam konfigurasi sosio-teknis tertentu.

Sintesis terhadap dimensi ODE menunjukkan bahwa *benign disinhibition* dan *toxic disinhibition* merupakan dua bentuk manifestasi yang memiliki konsekuensi berbeda. *Benign disinhibition* berkaitan dengan keterbukaan diri, ekspresi emosional, dan dukungan sosial, tetapi aspek tersebut masih relatif kurang dieksplorasi secara empiris. Sebaliknya, temuan penelitian lebih banyak terkonsentrasi pada *toxic disinhibition*, terutama dalam bentuk *cyberbullying*, *trolling*, agresi verbal, perilaku toksik, penyebaran hoaks,

pelanggaran privasi, dan ketidaksantunan komunikasi. Pola tersebut menunjukkan bahwa konsekuensi etis ODE dalam literatur lebih kuat terlihat pada persoalan pelemahan kontrol normatif dan akuntabilitas dalam interaksi digital. Dampaknya kemudian meluas dari relasi interpersonal menuju kualitas komunikasi dan ruang publik digital, termasuk potensi penurunan rasa aman, pelemahan norma komunikasi, serta gangguan terhadap kualitas deliberasi dan kohesi sosial.

Pada saat yang sama, penelitian ini menemukan bahwa pengetahuan mengenai ODE masih memiliki keterbatasan empiris, metodologis, dan kontekstual. Dominasi desain *cross-sectional*, penggunaan *self-report*, keterbatasan keragaman sampel, variasi instrumen, serta terbatasnya studi lintas budaya dan lintas platform membatasi kemampuan literatur untuk menjelaskan hubungan kausal dan variasi ODE pada konteks yang berbeda. Keterbatasan tersebut juga berhubungan dengan ketimpangan pemahaman antara *toxic* dan *benign disinhibition*. Karena itu, pemaknaan ODE dalam perspektif etika kewargaan digital perlu ditempatkan pada hubungan antara perilaku, tanggung jawab, dan struktur ruang digital. Temuan penelitian menunjukkan bahwa pendidikan kewarganegaraan perlu mengembangkan akuntabilitas digital, regulasi diri, kesadaran moral, dan kemampuan mempertimbangkan konsekuensi sosial dari tindakan daring, sementara platform memiliki tanggung jawab melalui desain dan tata kelola yang mendukung akuntabilitas serta interaksi yang sehat. Dengan demikian, kontribusi utama penelitian ini terletak pada pemaknaan ODE sebagai persoalan etika kewargaan dalam lingkungan sosio-teknis, sehingga respons terhadap disinsihibisi daring perlu bergerak secara

simultan pada kapasitas pengguna, pendidikan etika, norma sosial, dan tata kelola platform.

Penelitian selanjutnya perlu memperkuat bukti empiris mengenai ODE melalui desain longitudinal dan eksperimental untuk menguji arah hubungan antara karakteristik ruang digital, faktor psikologis, dan perilaku daring. Diversifikasi sampel ke berbagai kelompok usia, latar sosial-budaya, serta pengguna pada platform yang berbeda juga diperlukan untuk menguji karakter situasional ODE secara lebih kontekstual. Selain itu, penggunaan observasi perilaku digital, analisis data daring, dan instrumen pengukuran yang lebih terstandar dapat memperkuat validitas dan komparabilitas temuan. Kajian mendatang juga perlu memberikan perhatian lebih besar terhadap *benign disinhibition* agar pemahaman mengenai ODE tidak berhenti pada persoalan perilaku toksik, serta mengembangkan penelitian intervensi yang menguji efektivitas pendidikan regulasi diri, empati, kesadaran moral, dan etika kewargaan digital dalam mengurangi konsekuensi negatif ODE. Arah tersebut penting untuk menjembatani temuan mengenai ODE dengan pengembangan praktik pendidikan dan tata kelola ruang digital yang lebih bertanggung jawab.

Referensi

- Anggarani, F. K., & Amalia, F. (2021). Disinsihibisi Online sebagai Mediator Hubungan antara Kebingungan Identitas dan Cyberbullying pada Remaja. *Jurnal Psikologi Teori dan Terapan*, 11(2), 116. <https://doi.org/10.26740/jptt.v11n2.p116-127>

- Bulan, M. A. I. C., & Wulandari, P. Y. (2021). Pengaruh Kontrol Diri Terhadap Kecenderungan Perilaku Cyberbullying Pada Remaja Pengguna Media Sosial Anonim. *Buletin Riset Psikologi dan Kesehatan Mental (BRPKM)*, 1(1), 497–507.
<https://doi.org/10.20473/brpkm.v1i1.25127>
- Cheung, C. M. K., Wong, R. Y. M., & Chan, T. K. H. (2020). Online disinhibition: Conceptualization, measurement, and implications for online deviant behavior. *Industrial Management & Data Systems*, 121(1), 48–64.
<https://doi.org/10.1108/IMDS-08-2020-0509>
- Clark-Gordon, C. V., Bowman, N. D., Goodboy, A. K., & Wright, A. (2019). Anonymity and Online Self-Disclosure: A Meta-Analysis. *Communication Reports*, 32(2), 98–111.
<https://doi.org/10.1080/08934215.2019.1607516>
- Creswell, J. W. (2009). *Third Edition Research Design Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.
- Fakhri, N. (2025). *Digital Device Usage and Cyber-aggression: An Analysis of Online Aggressive Behavior Patterns on Digital Platforms in Indonesia*. 5(6).
- Febriana, S. K. T. & Fajrianthi. (2019). Cyber Incivility Perpetrator: The Influenced of Dissociative Anonymity, Invisibility, Asynchronicity, and Dissociative Imagination. *Journal of Physics: Conference Series*, 1175, 012238.
<https://doi.org/10.1088/1742-6596/1175/1/012238>
- Gan, W., Chen, Z., Wu, Z., Huang, X., & Wang, F. (2024). Aggression in online gaming: The role of online disinhibition, social dominance orientation, moral disengagement and gender traits among Chinese university students. *Frontiers in Public Health*, 12, 1459696.
<https://doi.org/10.3389/fpubh.2024.1459696>
- Gumelar, G., Maulana, H., Mas Bakar, R., & Erik, E. (2024). How online inhibition fuels incivility through moral disengagement. *Online Journal of Communication and Media Technologies*, 14(4), e202448.
<https://doi.org/10.30935/ojcm/15006>
- Haqie, D. A., Hapsari, W., & Karsiyati. (2024). *Peran Anonimitas Dan Konsep Diri Terhadap Online Disinhibition Effect Pada Remaja*.
<https://doi.org/10.5281/ZENODO.13761862>
- Julianti, A., Konradus, N., Susianti, D., & Maharani, A. (2025). Online Disinhibition Effect pada Anggota Komunitas Marah-marah di Platform X (Studi Deskriptif). *Arjwa: Jurnal Psikologi*, 4(3), 34–46.
<https://doi.org/10.35760/arjwa.2025.v4i3.14790>
- Juliati, K., Nugroho, S., & Qorifah, N. N. (2025). *Online Disinhibition and Anonymity in Adolescent Tiktok Discourse: Implication for Cyberbullying and Digital Education*. 5(2).
<https://doi.org/https://doi.org/10.51817/jpd.r.v5.i2.1357>
- Khairani, A., & Guspa, A. (2025). Pengaruh Self Control terhadap Toxic Online Disinhibition Effect pada Generasi Z Pengguna Aplikasi X. *CAUSALITA : Journal of Psychology*, 3(1), 243–253.

- <https://doi.org/10.62260/causalita.v3i1.525>
- Lai, C.-Y., & Tsai, C.-H. (2016). Cyberbullying in the Social Networking Sites: An Online Disinhibition Effect Perspective. *Proceedings of the The 3rd Multidisciplinary International Social Networks Conference on SocialInformatics 2016, Data Science 2016*, 1–6. <https://doi.org/10.1145/2955129.2955138>
- Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, 28(2), 434–443. <https://doi.org/10.1016/j.chb.2011.10.014>
- Lestari, D. R., & Alwi, M. A. (2024). Hubungan Online Disinhibition Effect Dengan Kecenderungan Perilaku Cyberbullying Pada Siswa MA Pengguna Instagram.
- Manalu, R. (2023). Fenomena Extreme Speech pada Ruang Virtual: Memahami Perilaku Ujaran Kasar di Media Sosial. *Jurnal Audience*, 6(1), 92–108. <https://doi.org/10.33633/ja.v6i1.8278>
- Martin, K. H. (2013). Leveraging Disinhibition to Increase Student Authority in Asynchronous Online Discussion. *Online Learning*, 17(3). <https://doi.org/10.24059/olj.v17i3.304>
- Matondang, A. F., & Afdal, A. (2023). The Relationship between Hoax Behavior and Toxic Disinhibition among Indonesian High School Students. *Journal of Educational, Health and Community Psychology*, 12(4), 822. <https://doi.org/10.12928/jehcp.v12i4.27288>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. <https://doi.org/10.1136/bmj.n71>
- Rafly, & Ningsih, Y. T. (2025). Pengaruh Dark Personality terhadap Online Disinhibition di Indonesia. *Social, Humanities, and Educational Studies*, 8(4), 866–872. <https://doi.org/https://doi.org/10.20961/shes.v8i4.110288>
- Ramadhani, A. Z., & Merida, S. C. (2024). Self-Control and the Phenomenon of Toxic Online Disinhibition in Teenagers Who Have Twitter. *Nusantara Journal of Behavioral and Social Sciences*, 3(2), 45–52. <https://doi.org/10.47679/202454>
- Samsiah, A. N., & Utami Sumaryanti, I. (2023). Pengaruh Anonimitas terhadap Perilaku Cyberbullying pada Remaja Akhir di Bandung. *Bandung Conference Series: Psychology Science*, 3(1). <https://doi.org/10.29313/bcps.v3i1.5362>
- Sánchez-Hernández, M. D., Herrera, M. C., & Expósito, F. (2024). Is Online Disinhibition Related to Cyberdating Abuse Perpetration through Moral Disengagement? The Moderating Role of Gender, Sexism, and Cybervictimization. *Sex Roles*, 90(7), 938–959. <https://doi.org/10.1007/s11199-024-01480-3>
- Shabira, A. M., & Lestari, A. F. (2024). Peran Mediasi Online Disinhibition Effect Dalam Pengaruh Fanatisme Terhadap Agresif Verbal (Studi Pada Komunitas ARMY Bogor Project). *Commptempor: Jurnal*

- Komunikasi Kontemporer*, 01(01), 66–81.
<https://doi.org/10.36782/cjik.v1i01.350>
- Siregar, F. M. O. (2026). *Pengaruh Online Disinhibition Effect terhadap Toxic Behavior pada Pemain Multiplayer Online Battle Arena Game* [Universitas Airlangga].
<http://repository.unair.ac.id/140065/>
- Soares, F. B., Gruzd, A., Jacobson, J., & Hodson, J. (2023). To troll or not to troll: Young adults' anti-social behaviour on social media. *PLOS ONE*, 18(5), e0284374. <https://doi.org/10.1371/journal.pone.0284374>
- Sudarja, K. (2025). Fenonema Kekerasan Verbal di Era Digital Sebagai Faktor Penurunan Kesantunan Pada Remaja. *ALFABETA: Jurnal Bahasa, Sastra, dan Pembelajaran*, 8(2).
<https://doi.org/10.33503/alfabeta.v8i2.2188>
- Suler, J. (2004). The Online Disinhibition Effect. *Cyberpsychology, Behavior, and Social Networking*, 7(3), 321–326. <https://doi.org/https://doi.org/10.1089/1094931041291295>
- Syrjämäki, A. H., Ilves, M., Olsson, T., Kiskola, J., Isokoski, P., Rantasila, A., Bente, G., & Surakka, V. (2024). Online disinhibition mediates the relationship between emotion regulation difficulties and uncivil communication. *Scientific Reports*, 14(1), 30019. <https://doi.org/10.1038/s41598-024-81086-7>
- Wang, C., Kue-Ing, & Shih, J.-F. (2014). Factors Influencing University Students' Online Disinhibition Behavior: The Moderating Effects of Deterrence and Social Identity. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 8. <https://doi.org/https://doi.org/10.5281/zenodo.1337205>
- Wang, L., Jiang, S., Zhou, Z., Fei, W., & Wang, W. (2024). Online disinhibition and adolescent cyberbullying: A systematic review. *Children and Youth Services Review*, 156, 1–9. <https://doi.org/https://doi.org/10.1016/j.chidyouth.2023.107352>
- Wu, B., Xiao, Y., Zhou, L., Li, F., & Liu, M. (2023). Why Individuals with Psychopathy and Moral Disengagement Are More Likely to Engage in Online Trolling? The Online Disinhibition Effect. *Journal of Psychopathology and Behavioral Assessment*, 45(2), 322–332. <https://doi.org/10.1007/s10862-023-10028-w>
- Wu, B., Zhou, L., Deng, Y., Zao, J., & Liu, M. (2022). Online Disinhibition and Online Trolling Among Chinese College Students: The Mediation of the Dark Triad and the Moderation of Gender. *Cyberpsychology, Behavior, and Social Networking*, 25(11), 744–751. <https://doi.org/https://doi.org/10.1089/cyber.2022.004>
- Wu, S., Lin, T.-C., & Shih, J.-F. (2017). Examining the antecedents of online disinhibition. *Information Technology & People*, 30(1), 189–209. <https://doi.org/10.1108/ITP-07-2015-0167>
- Yang, J., Wang, N., Gao, L., & Wang, X. (2021). Deviant peer affiliation and adolescents' cyberbullying perpetration: Online disinhibition and perceived social support as moderators. *Children and Youth Services Review*, 127, 106066.

<https://doi.org/10.1016/j.chilyouth.2021.106066>

Zahriyanto, A., Selo, G. K., Qodir, L. J., Zihni, A. A., Rabbani, G. A. N., & Apryono, C. B. I. K. (2025). Pancasila sebagai Fondasi Etika Sosial di Era Digital: Tantangan dan Peluang. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(4), 4782–4789.

<https://doi.org/10.31004/riggs.v4i4.4445>

Biarkan halaman ini tetap ada

[halaman ini sengaja dikosongkan]