

Bab 1 Pendahuluan

Latar Belakang

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam cara individu berinteraksi dan berkomunikasi. Hal tersebut didukung dengan adanya internet sebagai media yang memfasilitasi pertukaran informasi secara global dan *real time*. Kehadiran internet telah mengubah cara individu mengakses informasi, berbagi konten, dan membangun jaringan sosial yang lebih luas dari sebelumnya (Haris et al., 2024). Berbagai aktivitas seperti berdiskusi, bermain, belajar bahkan mencari pasangan kini dapat dilakukan dengan cara yang lebih menarik melalui internet (Rahim & Suryanto, 2018). Rahim & Suryanto (2018) pun dalam tulisannya menjelaskan bahwa dengan adanya internet, dapat membuat individu berkomunikasi dengan siapa pun, kapan pun, dan di mana pun tanpa khawatir dengan jarak dan waktu yang berbeda.

Asosiasi Penyelenggara Jasa Internet Indonesia (2024) telah merilis data terbaru hasil survei terhadap 8.720 responden yang disebarkan secara proporsional ke 38 provinsi yang ada di Indonesia. Data menunjukkan bahwa sebanyak 221.563.497 jiwa di Indonesia telah menggunakan internet, dari total populasi sebesar 278.696.200 jiwa yang berarti sekitar 79% masyarakat Indonesia telah terhubung ke internet. Angka ini menunjukkan bahwa lebih dari setengah populasi Indonesia telah menjadi bagian dari masyarakat digital.

Tingginya jumlah pengguna internet di Indonesia, tidak terlepas dari banyaknya layanan aplikasi atau konten di internet (*over the top*) yang dapat diakses secara mudah. Kehadiran *over the top* yang mempermudah cara berkomunikasi, berinteraksi, dan berbagi informasi, menandai peralihan dari metode tradisional ke digital berdasarkan Surat Edaran Nomor 3 Tahun 2016 terkait penyediaan layanan aplikasi dan/atau konten melalui internet (*over the top*). *Platform-platform* seperti Instagram, Youtube, Zoom dan WhatsApp telah menjadi bagian penting dalam kehidupan sehari-hari masyarakat (Like, 2021). Saat ini, media sosial merupakan salah satu layanan *over the top* yang banyak digunakan masyarakat di Indonesia. *Platform* ini menawarkan kemudahan dalam berkomunikasi secara daring, memudahkan penggunaannya untuk berbagi, menciptakan, serta berpartisipasi dalam berbagai bentuk konten (Sikumbang et al., 2024). Salah satu media sosial yang paling populer di Indonesia adalah Instagram, dengan jumlah lebih dari 85 juta atau 30,51% masyarakat Indonesia telah menggunakan Instagram dengan durasi penggunaan 1-2 jam perhari (APJII, 2024).

Dalam konteks ini, generasi Z menjadi kelompok yang paling terpengaruh dengan kemajuan teknologi. Generasi Z mencakup individu yang lahir antara tahun 1997 hingga 2012

(Abravanel, 2020; BPS, 2020) adalah generasi pertama yang mengalami masa pertumbuhan dengan akses luas terhadap teknologi digital dan media sosial (Alferjany & Alias, 2020). Tak heran jika generasi Z sudah sangat familiar dengan *smartphone*, media sosial, dan berbagai teknologi lainnya. Mereka yang berusia 18-24 tahun menjadi salah satu kelompok yang banyak mengakses Instagram, sebesar 32,7% dari total 90.183.200 pengguna Instagram di Indonesia saat ini (NapoleonCat, 2024). Dari data tersebut artinya Generasi Z merupakan kelompok usia yang rentan terpengaruh oleh arus global dan dampak negatif dari internet termasuk dari sosial media Instagram (Abdillah & Handoko Putro, 2022).

Instagram sering kali menjadi *platform* utama bagi generasi Z untuk berinteraksi, berbagai konten, dan membangun eksistensi diri. Namun di sisi lain, Instagram juga menyediakan ruang bagi mereka untuk mengekspresikan agresi secara terbuka, melalui komentar negatif, kritik tajam, atau bahkan ancaman (Sobkin & Fedotova, 2021). Instagram berperan sebagai medium penularan emosi, terutama kemarahan, karena emosi yang tinggi dapat memperkuat jaringan sosial dan menyebarkan informasi dengan cepat (Fang et al., 2018). Kemarahan ini kadang-kadang disebarkan oleh individu yang tidak dikenal secara langsung, yang bersama-sama membentuk perilaku digital dan memengaruhi opini publik. Salah satu masalah yang muncul dari pengguna Instagram adalah maraknya tindakan penindasan *online*, atau yang dikenal sebagai *cyberbullying* (Santika & Krisnayana, 2022). Lalu Instagram juga dapat menyediakan akses yang mudah bagi banyak potensi negatif seperti ujaran kebencian, penipuan, pornografi, penculikan dan lain sebagainya (Nugraha et al., 2023).

Salah satu ruang digital di Instagram yang memperlihatkan intensitas interaksi negatif dapat ditemukan di akun publik @lambe_turah dengan 12,3 juta akun yang mengikuti dan lebih dari 19,1 ribu postingan (06/01/2025). Akun ini menarik perhatian banyak pengguna, termasuk kalangan generasi Z yang dikenal aktif mengikuti isu-isu viral dan cenderung responsif terhadap konten yang sedang banyak diperbincangkan publik. Sebuah unggahan di akun ini bisa meraih hingga 17,9 juta tayangan, 289 ribu tanda suka, 10,7 ribu komentar dan 16,1 ribu kali dibagikan. Namun, dibalik tingginya *engagement* di akun tersebut, muncul pola interaksi yang cenderung agresif dan tidak terkontrol, terutama dalam kolom komentar. Kolom komentar kerap dipenuhi kata-kata kasar, ujaran bernada sinis, hinaan, hingga serangan personal, baik antar pengguna maupun terhadap objek dalam unggahan akun tersebut (Agustina, 2021). Konten yang diunggah pun tidak jarang membentuk opini publik dengan cara yang provokatif, memancing emosi, dan menurunkan batasan etika dalam

berinteraksi (Tyaswanti et al., 2024). Generasi Z, sebagai kelompok pengguna yang dominan dan sangat aktif di Instagram, menjadi salah satu yang paling rentan menampilkan pola interaksi semacam ini karena kedekatannya dengan teknologi serta kecenderungan untuk membentuk identitas diri secara digital (Afifah & Kuntari, 2025; Sihombing, 2024).

Meskipun Instagram memberikan banyak fasilitas dan kemudahan penggunaannya sebagai alat komunikasi ternyata tidak menghasilkan interaksi yang autentik seperti halnya komunikasi tatap muka (Nugraha et al., 2023). Dalam kenyataannya dapat membuat seseorang yang sehari-harinya dikenal sebagai pribadi yang baik dan tidak mudah melakukan perilaku negatif, di Instagram mereka berani berkomentar kasar, menghina, memberikan ancaman, bahkan tak segan untuk membagikan konten sensitif ke publik. Hal ini menegaskan bahwa banyak pengguna yang lebih terbuka di Instagram dan sering berinteraksi tanpa batasan, namun dalam kehidupan nyata, mereka cenderung tertutup dan enggan bersosialisasi (Rahayu et al., 2019).

Fenomena pada paragraf di atas sangat lekat terjadi dalam kehidupan sehari-hari pengguna Instagram di Indonesia. Bila merujuk pada hasil survei *World Population Review* (2024) yang menunjukkan negara Indonesia menempati urutan ke-4 sebagai negara paling ramah di dunia, dengan etika dan kesopanan yang dinilai sangat tinggi. Namun, dalam dunia digital, perilakunya justru sangat berbeda. Dari laman berita CNBC Indonesia yang dilaporkan oleh Bestari (2023) disebutkan bahwa Ridwan Kamil (selaku Gubernur Jawa Barat) mengkritik pengguna sosial media atau *netizen* di Indonesia yang kerap kali berperilaku *julid* dan *ngerujak* dengan memberikan komentar negatif di media sosial. Bahkan ia mengakui bahwa tidak heran jika netizen di Indonesia menjadi juara terkasar di Asia Fasifik. Hal ini sesuai dengan laporan Microsoft "*Digital Civility Index (DCI)*" yang dituliskan oleh Mazrieva (2021) dimana survei terhadap 16.000 responden yang tersebar di 32 negara pada tahun 2020 menunjukkan bahwa Indonesia merupakan negara yang menempati urutan ke-29, dengan kata lain *netizen* Indonesia memiliki tingkat kesopanan *online* terendah di Asia Tenggara. Survei Microsoft mengidentifikasi tiga risiko utama di dunia digital yang dihadapi oleh warga Indonesia, yaitu penipuan dan hoaks, ujaran kebencian, serta diskriminasi (Pusparisa, 2021).

Dalam ranah Psikologi, fenomena bedanya perilaku masyarakat yang ditampilkan dalam dunia *online* dan *offline* ini disebut sebagai *online disinhibition* (A. Joinson, 1999; A. N. Joinson, 2001, 2007; Suler, 2004). *Online disinhibition effect* memungkinkan individu untuk mengekspresikan perasaan atau keinginan yang biasanya tidak mereka tunjukkan dalam

kehidupan sehari-hari. Kemudian terlebih lagi Instagram sendiri menyediakan akses yang mudah bagi banyak potensi *negative*, banyak masyarakat pada akhirnya terlibat perilaku negatif seperti *harassment*, *cyberbullying* bahkan pornografi yang sebenarnya sangat beresiko jika dilakukan secara langsung. Hal ini mengindikasikan suatu fenomena spesifik yang saat ini marak terjadi di media sosial yaitu *toxic disinhibition online* (Suler, 2004).

Menurut A. N. Joinson (2007) *toxic disinhibition online* merupakan kecenderungan individu untuk menunjukkan perilaku negatif yang berbeda dari interaksi *offline*, karena berkurangnya kekhawatiran terhadap penilaian sosial dan cara menampilkan diri. Lebih lanjut, A. N. Joinson (2001) menyebutkan bahwa *toxic disinhibition online* ini dipengaruhi oleh identitas personal yang ditunjukkan individu dalam interaksinya dengan orang lain pada saat *online*, yang dapat dilakukan dengan keterbukaan diri (*self-disclosure*). Namun dalam konteks perilaku, *toxic disinhibition online* sering kali menunjukkan tindakan agresif seperti *online flaming*, di mana individu menggunakan *platform online* untuk mengekspresikan kemarahan atau kritik yang tajam dengan cara yang mungkin tidak mereka lakukan dalam kehidupan nyata (Lapidot-Lefler & Barak, 2012). Perilaku yang dimaksud dapat berupa penggunaan kata-kata kasar, kritik yang tidak konstruktif, ujaran kebencian, ancaman, dan bahkan bisa berujung pada *cyberbullying* (Kiswanto et al., 2022). Hal ini sering terjadi karena individu cenderung abai terhadap aturan sosial dan hambatan yang ada dalam interaksi langsung. Sejumlah data menunjukkan adanya pola perilaku yang sering ditemukan di Instagram diantaranya: agresivitas, *cyberbullying*, *harassment*, dan penyebaran konten tidak pantas, termasuk pornografi, memperkuat adanya pola perilaku yang cenderung destruktif ini (Laora & Sanjaya, 2021).

Kajian-kajian yang menunjukkan data terbaru mengenai *toxic disinhibition online* di Indonesia sudah tercatat. Terdapat sebuah penelitian yang menunjukkan bahwa perilaku *toxic disinhibition online* di Instagram terutama di kalangan generasi Z, berkaitan erat dengan maraknya kasus *cyberbullying*. Survei yang dilakukan oleh *Center for Digital Society* (CfDS) dan *Center for Lifespan and Development* (CLSD) Universitas Gadjah Mada (UGM) menemukan bahwa 1.182 dari 3.077 responden, atau 45,35%, mengaku pernah mengalami perundungan di Instagram. Sementara itu, 45,35% responden lainnya juga melaporkan bahwa mereka pernah menjadi korban (Sucahyo, 2021). Selain itu, kasus *hate speech* atau ujaran kebencian juga banyak ditemukan di Instagram selama masa pemilu 2024, di mana lebih dari 24.000 teks ujaran kebencian dari 4.472 konten ujaran kebencian yang berhasil dipantau dengan jumlah komentar sebanyak 9 juta dan like sebanyak 181 juta selama periode 1

September 2023 sampai 31 Januari 2024 (Saptoyo & Erdianto, 2024). Hal ini ditambah dengan tingginya *engagement* terhadap postingan-postingan yang mengandung ujaran kebencian.

Meskipun Instagram telah menghadirkan fitur *sensitive content control* untuk membatasi akses konten negatif, penelitian sebelumnya menunjukkan bahwa masih banyak pengguna Instagram terutama generasi Z, tetap bisa mengakses konten beresiko seperti pornografi, agresivitas dan konten kekerasan lainnya (Riyanto & Nistanto, 2021). Tentunya hal tersebut dapat semakin memperparah efek *toxic online disinhibition*, dimana minimnya hambatan yang dirasakan individu untuk mengekspresikan emosi negatif, kritik tajam, dan perilaku agresif tanpa mempertimbangkan konsekuensinya di Instagram dapat memicu stres, kecemasan, dan depresi (Vaterlaus et al., 2021). Jika diamati lebih jauh, *toxic online disinhibition* dapat terjadi karena pengguna Instagram merasa kurang terikat oleh norma sosial yang biasanya mengatur interaksi langsung, mereka cenderung terlibat dalam perilaku merusak yang dapat memperburuk kondisi mental mereka sendiri dan orang lain (Suler, 2004).

Maraknya perilaku *negative netizen* di Instagram menandakan bahwa fenomena *toxic disinhibition online* sudah menjadi masalah serius di Indonesia. Perilaku seperti meluapkan emosi berlebihan, menggunakan bahasa kasar, ujaran kebencian, ancaman, mengakses konten sensitif dan tindakan agresif lainnya dapat menyebabkan konflik, trauma, atau bahkan ide bunuh diri pada korban (Rizqi, 2023). Penelitian sebelumnya dari Nurgraha dkk pada tahun 2023 menyarankan kepada peneliti selanjutnya untuk melibatkan kelompok populasi yang lebih luas guna mendapatkan gambaran yang lebih komprehensif tentang fenomena *toxic disinhibition online*, serta menggunakan platform *online* yang berbeda (Nugraha et al., 2023). Berdasarkan saran tersebut, untuk membuktikan adanya fenomena *toxic disinhibition online*, peneliti melakukan studi awal pada generasi Z yang merupakan pengikut akun @lambe_turah di Instagram. Peneliti menyebarkan kuesioner *online* melalui *google formulir* kepada 30 orang yang termasuk kategori generasi Z. Dari total 30 orang generasi Z yang berpartisipasi diketahui 10 orang berjenis kelamin laki-laki dan 20 orang perempuan.

Dari hasil studi awal yang dilakukan, peneliti menemukan bahwa mayoritas responden telah menggunakan Instagram lebih dari tiga tahun, dengan durasi penggunaan 2-3 jam perhari. Sebanyak 53,3% responden menyatakan kepemilikan akun Instagram lebih dari satu dengan identitas palsu atau anonim, dan 26,7% responden memiliki lebih dari satu akun Instagram dengan salah satu akunya beridentitas asli dan satunya lagi akun dengan identitas

anonim, serta 20% responden hanya memiliki satu akun Instagram dengan identitas anonim. Selain itu, mayoritas responden telah mengikuti akun @lambe_turah selama > 6 bulan. Kemudian peneliti juga mengukur tingkat *toxic online disinhibition* dengan menggunakan skala *toxic online disinhibition* yang diadaptasi oleh Addauliyah (2022).

Tabel 1.1 Gambaran kecenderungan toxic online disinhibition pada generasi Z di Instagram

No.	Item	STS	TS	S	SS
1.	Saya terkadang merasa seperti orang yang berbeda saat berinteraksi di Instagram.	0%	16,7%	56,7%	26,7%
2.	Lebih mudah bagi saya berkomunikasi di Instagram, karena saya dapat menjawab kapan saja saya suka.	10%	16,7%	53,3%	20%
3.	Tidak ada aturan dalam berkomunikasi di Instagram sehingga saya dapat melakukan apapun yang saya inginkan.	26,7%	20%	46,7%	6,7%
4.	Saya merasa tidak keberatan menuliskan kata-kata kasar di Instagram, karena bersifat anonim. (contoh: mengetik kata hewan yang ditujukan untuk orang lain, anjir, dll)	3,3 %	20%	56,7%	20%
5.	Lebih mudah bagi saya menuliskan di Instagram untuk hal-hal yang sulit dikatakan di kehidupan nyata karena saya tidak melihat wajah orang lain.	6,7%	26,7%	40%	26,7%
6.	Menurut saya lebih mudah untuk menuliskan hal-hal yang menghina orang lain di Instagram, karena tidak adanya dampak yang signifikan.	26,7%	20%	33,3%	20%

***Keterangan:** STS =Sangat tidak sesuai; S = Sesuai; S=Sesuai; SS= Sangat sesuai

Berdasarkan tabel 1.1 diperoleh hasil bahwa dari semua jawaban responden rata-rata memilih sesuai sebagai jawaban atas apa yang mereka alami saat ini. Merasa seperti orang yang berbeda saat berinteraksi di Instagram dan merasa tidak keberatan menuliskan kata-kata kasar di Instagram, karena bersifat anonim menjadi jawaban yang paling sesuai dengan perilaku mereka ketika berinteraksi di Instagram yaitu sebesar 56,7%. Mereka juga merasa

lebih mudah berkomunikasi di Instagram, karena dapat menjawab kapan saja mereka suka (53,3%). Lalu sebanyak 46,7% responden merasa bebas melakukan apapun yang mereka inginkan karena menganggap tidak ada aturan dalam berkomunikasi di Instagram. Selanjutnya 40% dari responden menganggap lebih mudah menuliskan hal-hal yang sebelumnya sulit dikatakan pada saat *offline* karena mereka tidak melihat wajah orang lain secara langsung. Terakhir sebesar 33,3% menganggap tidak ada dampak yang signifikan sehingga mereka tidak segan untuk menghina orang lain di Instagram.

Berdasarkan hasil studi awal yang diuraikan di atas, membuktikan bahwa fenomena *toxic disinhibition online* ini telah menjadi masalah yang benar-benar terjadi pada generasi Z. Generasi Z disini memiliki kecenderungan berperilaku *toxic disinhibition online* saat berinteraksi di salah satu ruang digital akun @lambe_turah di Instagram, sehingga generasi ini dipilih sebagai subjek untuk melakukan analisis lebih mendalam mengenai *toxic disinhibition online*. Kemudian, berdasarkan studi awal di atas, ditemukan bahwa sebagian besar responden menyebutkan bahwa anonimitas sangat mempengaruhi perilaku *toxic disinhibition online*. Hampir seluruh responden mengaku bahwa saat berinteraksi di akun @lambe_turah mereka merasa tidak keberatan menuliskan kata-kata kasar di Instagram karena bersifat anonim.

Selanjutnya guna memperdalam isi dari respon yang diberikan responden, maka peneliti juga melakukan wawancara kepada SNZ salah satu responden yang mengaku setiap item yang disajikan sangat sesuai dengan perilakunya saat ini.

“Dalam interaksi tatap muka, saya mungkin merasa lebih sadar akan bagaimana saya dipersepsikan oleh orang lain, yang dapat menghambat saya untuk mengekspresikan diri secara bebas. Di Instagram, saya tidak merasa ada tekanan dari orang lain, sehingga mungkin merasa lebih bebas untuk bertindak. Di Instagram @lambe_turah saya pernah mengomentari postingan kasus artis karena perselingkuhan. Saya juga pernah ikut ngebully Fuji karena waktu itu kesal dengan perlakuan fansnya yang begitu fanatik”.

Dalam wawancara bersama SNZ tersebut diketahui bahwa telah terjadi perilaku *toxic disinhibition online* pada generasi Z, dengan penyebab utamanya yaitu keberadaan akun anonim, seperti yang diungkapkan:

“Akun saya tidak memakai foto asli dan nama asli, sehingga identitas saya tidak diketahui atau anonim. Jadi saya merasa lebih berani dan lebih bebas untuk berkomentar di Instagram. Saya menggunakan akun berbeda karena agar tidak diketahui orang yang dikenal”.

Berdasarkan hasil studi awal di atas dapat teridentifikasi bahwa *toxic disinhibition*

online yang terjadi pada generasi Z didominasi oleh pengguna akun anonim. Peneliti juga menemukan bahwa terdapat dua kepentingan yang berbeda dari penggunaan Instagram oleh generasi Z yang memiliki lebih dari satu akun atau mempunyai *second account* atau akun kedua. Mayoritas generasi Z yang mengikuti akun @lambe_turah mempunyai akun kedua menggunakan identitas yang bukan sebenarnya (anonim) atau berbeda dengan akun pertamanya. Dengan kata lain bagi mereka yang berinteraksi di Instagram menggunakan akun anonim, tidak segan melakukan perilaku kurang baik di Instagram.

Suler (2004) dalam tulisannya menyebutkan ada enam faktor yang dapat mempengaruhi *toxic online disinhibition* yaitu *dissociative anonymity*, *invisibillity*, *asynchronicity*, *solipsistic interjection*, *dissociative imagination*, dan *minimization of status and authority*. Selain itu, Karima & Adinia (2019) membuat simpulan dari beberapa literatur dalam dan luar negeri terdapat beberapa faktor yang dapat menyebabkan terjadinya *toxic online disinhibition*. Faktor-faktor ini meliputi lingkungan internet yang mengurangi interaksi tatap muka, aspek psikologis yang berkaitan dengan konsep diri dan kegagalan dalam mengontrol diri, serta faktor eksternal seperti kemampuan untuk mengendalikan situasi, identitas sosial, dan kepastian hukuman. Berdasarkan uraian faktor-faktor di atas, faktor internal seperti anonimitas dan faktor psikologis menjadi faktor penentu dan paling dominan di antara yang lainnya (Gackenbach, 2007)

Dalam ilmu psikologi, kondisi dimana identitas seorang individu tidak dapat teridentifikasi oleh individu lainnya atau yang sering diartikan sebagai anonimitas (Lee et al., 2014). Suler (2024) mengidentifikasi bahwa anonimitas memungkinkan seseorang untuk mengekspresikan diri secara lebih bebas, yang dapat mengarah pada perilaku yang lebih terbuka, baik positif maupun negatif. Anonimitas ini meliputi tiga aspek utama yaitu *unlikability*, *unobservability*, dan *pseudonyity* anonimitas (Lee et al., 2014). Dari *literatur review* yang dilakukan oleh peneliti, ditemukan bahwa anonimitas merupakan *predictor* penting dalam memprediksi bagaimana seorang individu berperilaku dalam konteks *online*.

Seperti penelitian yang dilakukan oleh Lapidot-Lefler & Barak (2012) yang menemukan bahwa anonimitas dapat meningkatkan perilaku agresif dalam interaksi *online*. Individu menunjukkan bahwa ketika identitas pengguna disembunyikan, kemungkinan untuk terlibat dalam perilaku merugikan seperti *trolling* dan *cyberbullying* meningkat. Selanjutnya penelitian yang dilakukan oleh Silke (2003) mendukung temuan lainnya yaitu dengan menunjukkan bahwa individu yang anonim cenderung menunjukkan perilaku yang lebih agresif dibandingkan dengan mereka yang teridentifikasi, memungkinkan untuk perilaku

lebih kasar dan kurang bertanggung jawab dalam pengalaman *online*. Sebaliknya, ketika individu berpartisipasi dalam lingkungan *online* dengan identitas yang diketahui, perilaku agresif berkurang dan perilaku prososial meningkat karena kesadaran bahwa tindakan mereka dapat dilacak kembali ke mereka secara pribadi (Christopherson, 2007). Sejalan dengan pendapat tersebut Lapidot-Lefler & Barak (2012) menemukan bahwa interaksi *online* tanpa anonimitas cenderung lebih konstruktif dan kurang merugikan. Hasil dari beberapa penelitian terdahulu pun telah membuktikan bahwa anonimitas mempunyai hubungan secara positif terhadap *toxic disinhibition online* (Rizqi, 2023). Dalam penelitiannya Rizqi (2023) menyatakan bahwa semakin tinggi anonimitas di sosial media cenderung akan meningkatkan perilaku *toxic disinhibisi online* pada penggunaanya.

Berbeda dengan hasil-hasil penelitian sebelumnya, penelitian lainnya menemukan bahwa anonimitas tidak selalu mempengaruhi perilaku *online* dengan cara yang sama. Berdasarkan studi Christopherson (2007) menunjukkan bahwa anonimitas memberikan kesempatan positif bagi individu, khususnya kaum muda, untuk mengeksplorasi dan mengembangkan identitas pribadi mereka tanpa takut akan penilaian negatif atau konsekuensi sosial. Hal tersebut didukung dengan hasil penelitian yang menyatakan bahwa terdapat pengaruh yang signifikan antara anonimitas dan *online disinhibition* terhadap kesehatan mental positif (Hanifah & Rozi, 2022). Perbedaan hasil yang ditemukan di atas menunjukkan adanya inkonsistensi atau gap penelitian yang diakibatkan oleh waktu dan responden penelitian yang berbeda dalam kurun waktu yang cukup lama.

Selain anonimitas, dalam berinteraksi secara *online* kontrol diri juga memiliki hubungan yang erat kaitannya. Suler (2004) menjelaskan bahwa "*disinhibition*" adalah suatu bentuk dari kontrol diri yang gagal. Kontrol diri merupakan kemampuan individu untuk mengesampingkan atau menyesuaikan dorongan (impuls) dalam diri, serta menginisiasi perilaku, pikiran, dan emosi guna mencapai tujuan jangka panjang (*goal-directed behavior*) (D. T. D. De Ridder et al., 2011). Kontrol diri berperan penting dalam menentukan apakah individu akan menunjukkan perilaku toxic atau tidak, karena kontrol diri merupakan kemampuan individu dalam mengendalikan dirinya ketika berinteraksi dengan orang lain. Individu dengan kontrol diri yang tinggi biasanya mampu menahan diri dari berbagai dorongan untuk berperilaku toxic, sedangkan mereka dengan kontrol dirinya yang rendah cenderung tidak mampu menahan diri dan beresiko mengalami *ego depletion* atau penurunan kemampuan diri dalam mengatur emosi, perilaku, dan mengambil keputusan (Ramadhani & Merida, 2024). Mereka dengan kontrol diri yang rendah mungkin tidak lagi mampu

merespons secara positif terhadap situasi yang memerlukan kontrol, sehingga rentan untuk terlibat dalam perilaku agresif (Erpiyani, 2021). Individu yang memiliki perilaku agresif biasanya menyerang objek dengan kontrol emosional yang rendah, akibatnya berpuncak pada manifestasi perilaku *toxic disinhibition online*. Pentingnya kontrol diri terlihat dalam studi awal penelitian ini, berdasarkan hasil wawancara bersama SNZ mengungkapkan:

“Kalau lagi di Instagram, buat saya penting banget untuk jaga kontrol diri. Biasanya kalau udah kesel sama orang, saya lebih milih untuk diam, tarik nafas, atau bahkan log out dulu, daripada kepancing dan ikutan emosi kan. Kontrol diri itu ibarat rem yang bikin kita gaasal ngomong atau posting sesuatu yang negative. Saya sendiri berusaha nggak ikut-ikutan toxic ya apalagi pas liat postingan-postingan di @lambe_turah. Kadang susah sih apalagi kalau emosi atau nggak suka sama konten tertentu, tapi kalau kita nggak bisa kontrol diri, ujung-ujungnya hanya akan membuat masalah untuk diri sendiri”.

Penelitian Nugroho (2022) memaparkan hasil temuannya terhadap 154 orang responden yaitu pemain aktif *game online* berusia 18-25 tahun yang teridentifikasi *toxic disinhibition online*, menunjukkan adanya korelasi *negative* terhadap kontrol diri. Artinya, semakin kuat kontrol diri yang dimiliki individu maka semakin rendah kecenderungan terhadap *toxic disinhibition online*, sebaliknya semakin rendah kontrol diri individu akan semakin tinggi kecenderungan terhadap *toxic disinhibition online*. Hal ini sejalan dengan penelitian Ramadhani & Merida (2024) yang memperoleh hasil yang sama yaitu terdapat hubungan negatif signifikan antara kontrol diri dan *toxic disinhibition online* pada remaja pengguna Twitter, menandakan bahwa kontrol diri mampu menjadi salah satu predictor *toxic disinhibition online*. Kemudian temuan ini diperkuat dengan adanya hasil penelitian terdahulu yang membuktikan bahwa kontrol diri berpengaruh secara signifikan terhadap perilaku *disinhibition online effect* (Arsha, 2022).

Berbeda dari penelitian-penelitian sebelumnya di atas, Nugraha et al. (2023) dengan penelitiannya yang dilakukan pada 154 mahasiswa menemukan bahwa, lebih dari setengah sampel (50,6%) yang diteliti mempunyai kontrol diri yang tinggi dengan hasil uji statistik memperlihatkan skor signifikansi 0,604 lebih besar dari 0,05. Sehingga kesimpulannya dalam penelitian tersebut kontrol diri tidak berhubungan dengan *toxic disinhibition online*. Perbedaan ini dapat disebabkan oleh penggunaan metode penelitian yang berbeda seperti penggunaan instrumen dan pemilihan subjek penelitian yang berbeda. Penelitian, Nugraha et al. (2023) menggunakan sampel mahasiswa dan menemukan bahwa kontrol diri tidak mempunyai hubungan atau dampak terhadap perilaku *toxic disinhibition online*, berbeda

dengan hasil penelitian Nugroho (2022) dan Ramadhani & Merida (2024) yang menunjukkan korelasi negatif signifikan antara kedua variabel tersebut. Beberapa hasil penelitian di atas mengindikasikan adanya pertentangan dalam penelitian lain yang dibahas dalam tulisan ini.

Mengingat besarnya risiko yang muncul secara signifikan akhir-akhir ini dari berbagai interaksi *online* yang secara langsung dapat mengarah pada fenomena “*toxic disinhibition online*”, terutama di kalangan generasi Z yang menjadi kelompok terbanyak pengguna jejaring media Instagram. Fenomena ini seringkali mengakibatkan berbagai perilaku negatif yang tidak sesuai dengan norma sosial atau merugikan dalam konteks *online*. Meskipun sudah ada berbagai fitur yang disediakan platform digital dan upaya lainnya untuk membatasi perilaku *toxic disinhibition* tersebut, namun kenyataannya masih banyak generasi Z pengguna Instagram yang terlibat dalam perilaku merugikan seperti *bullying*, ujaran kebencian, dan agresivitas di media sosial. Oleh karena itu, di era serba digital ini, sangat diperlukan penelitian yang lebih mendalam tentang *toxic disinhibition online*.

Penelitian ini tidak hanya penting, tetapi juga memberi kebaruan yang baik dalam memahami perilaku pengguna Instagram di kalangan generasi Z. Penelitian ini mengembangkan model konseptual dengan mengkaji kontrol diri sebagai variabel moderator dari pengaruh anonimitas terhadap *toxic disinhibition online*, berdasarkan saran penelitian sebelumnya (Rizqi, 2023). Tentunya penelitian ini mengisi kekosongan dalam literatur yang ada dan belum pernah dilakukan dalam penelitian *toxic disinhibition online* (Arsha, 2022; Nugraha et al., 2023; Nugroho, 2022; Ramadhani & Merida, 2024). Target penelitian yaitu penggunaan Instagram pun memberikan persepektif baru yang relevan dengan dinamika sosial generasi Z saat ini, mengingat fenomena perilaku negatif yang terus meningkat di platform media sosial Instagram.

Berdasarkan fenomena dan masalah yang telah dipaparkan di atas, serta ditemukannya inkonsistensi dari penelitian sebelumnya, serta saran dari penelitian sebelumnya menjadi pendorong bagi peneliti dalam melakukan penelitian mengenai “*Pengaruh anonimitas terhadap toxic disinhibition online dengan kontrol diri sebagai variabel moderator pada generasi Z di Instagram*”.

Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, masalah yang dapat dirumuskan dalam penelitian ini adalah:

1. Apakah kontrol diri berperan sebagai variabel moderator pada pengaruh anonimitas terhadap *toxic disinhibition online* pada generasi Z pengguna di Instagram?

2. Apakah anonimitas berpengaruh terhadap *toxic disinhibition online* pada generasi Z di Instagram?
3. Apakah kontrol diri berpengaruh terhadap *toxic disinhibition online* pada generasi Z di Instagram?

Tujuan Penelitian

Adapun tujuan dalam penelitian ini yaitu sebagai berikut:

1. Untuk mengetahui peran kontrol diri sebagai moderator perngaruh anonimitas terhadap *toxic disinhibition online* pada generasi Z di Instagram.
2. Untuk mengetahui pengaruh anonimitas terhadap *toxic disinhibition online* pada generasi Z di Instagram.
3. Untuk mengetahui pengaruh kontrol diri terhadap *toxic disinhibition online* pada generasi Z di Instagram.

Kegunaan Penelitian

Kegunaan teoritis

Penelitian ini diharapkan dapat memberikan wawasan bagi para pembaca terkait psikologi sosial, terutama dalam memahami dinamika perilaku *online* di media sosial. Hasil penelitian ini diharapkan dapat memperkaya literatur mengenai fenomena *toxic disinhibition online*, anonimitas, dan kontrol diri dalam lingkup sosial.

Kegunaan praktis

Secara praktis, hasil penelitian ini diharapkan dapat digunakan sebagai acuan bagi platform media sosial dalam mengembangkan kebijakan fitur yang lebih efektif untuk mengurangi perilaku *toxic disinhibition online*. Selain itu, hasil penelitian ini dapat memantik studi lanjutan, terutama dalam merancang intervensi yang dapat mengurangi *toxic disinhibition online*, dengan fokus pada pengaruh kontrol diri yang baik dan pengelolaan anonimitas di media sosial.