Article

Negative Prompts and the Affective Economies of Failure in Visual Generative AI Systems

CHEERU PADMA, Pondicherry University, India; e-mail: cheerulistens@pondiuni.ac.in
M SHUAIB MOHAMED HANEEF, Pondicherry University, India; e-mail: shuji22@pondiuni.ac.in

10.2478/bsmr-2025-0003

ABSTRACT

Every negative prompt is an algorithmic 'thou shalt not,' an act of negating and negotiating what "we do not wish to see" in a generative media output. This study identifies negative prompts in text-to-image generative systems as a crucial site where the digital error discourse becomes materially realised. Employing a multimodal digital ethnographic approach, this paper seeks to understand how these prompts extend beyond their instrumental use to see how they are entangled with the notions of failure and repair in AI systems. Through close consideration of this tool, the paper contends that generative systems sustain user engagement through an affective economy of failure: where errors are rendered ordinary and becomes the burden of users while a relentless update culture naturalises the logic of 'versions-as-evolutions,' ultimately cultivating a sort of cruel optimism among its users.

KEYWORDS: generative AI; negative prompts; affective economies; algorithmic failure; computational creativity

INTRODUCTION

> When the magician pulls
> the rabbit from the hat, the
> spectator can respond either with
> mystification or with curiosity.
> He can enjoy the surprise and
> the wonder of the unexplained
> (and perhaps inexplicable), or
> he can search for an explanation.
> (Simon & Newell, 1971)

This is how Simon and Newell set about their essay "Human Problem Solving: The State of the Theory in 1970," one of the seminal and formative works on what we call 'Artificial Intelligence' today. Fifty-five years later, it appears that the same analogy can help us make sense of the time we live in – when Generative Artificial Intelligence (Gen AI, Generative AI, or Generative systems hereafter) has become the zeitgeist. Because this magic, which Chesher and Albarrán-Torres call 'autolography'[1] (Chesher & Albarrán-Torres, 2023), has become a somewhat mundane part of our lives today, with poems, people, parallel universes and of course 'rabbits' – eerily lifelike – being conjured up within a matter of seconds using Generative AI. And the same is what it takes to make these vanish as well – a little prompt whispering in a generative system of your choice. But do we, the audience, even know his prop here, let alone the trick? Because, be it hat, hammer or 'hypercreative' AI, they are

1 Coined from the Greek words automatos, logos and graphos which roughly translates to 'self word drawing.'

all, in Heideggerian terms, tools that are ready to hand[2] – they fade into the background while quietly shaping our everyday practices. In addition, the AI tool, much like the magician's hat that hides the sleight of hand, remains an opaque and closed system – a black box, as it is called – with just an 'algorithmic aperture' (Amoore, 2020) that sees, draws in, discards and churns out data points. Yet this readiness to hand, according to Heidegger, lasts only until they seem to work 'seamlessly' in the background. What happens when the magic fails, when these systems break? The ruptures produced thereafter make these tools be present at hand,[3] for comprehension, for us to see it for something more than its use value. No longer sustained by the (mis) perceived ease of magic, we turn to the tool and the act, only to realise that bridging the disjunction between the act of magic and its intended effect requires considerable effort and is far from costless (Larsson & Viktorelius, 2024). This paper is a search for explanation in the ruptures, an attempt to demystify the (un)magical 'mistakes' and reparative revisions that contemporary visual Gen AI systems put forward, and the kind of affective economy that it fortifies. To ground this discussion, we consider the case of negative prompts (--no parameter in the case of Midjourney) to try to look beyond the 'what for' in its use and discourse.

The paper is divided into three sections. The first section is mostly aimed at outlining the key conceptual frameworks. Here, we situate our use of the term generative AI within the context of existing scholarship and briefly delineate the specific subset of text-to-image systems and negative prompts which form the empirical focus of our study. We end this part by engaging with literature on technological failure, which sets the stage for further discussions on its affective dimensions. The second section examines how generative AI industries actively cultivate an affective economy of failure, in which negative prompts recast AI artists as agents of maintenance and repair. Finally, in the third part, we theorise these observations, substantially drawing on Appadurai and Alexander's framework of failure as an affective economy.

2 'Ready to hand,' according to Heidegger, is the mode of being of a tool where it is encountered as a useful equipment, transparently integrated into an activity and not as an object of conscious reflection.

3 'Present at hand,' *per* Heidegger, is the mode of encountering entities as objectively present objects for theoretical observation, which occurs when a tool breaks.

GENERATIVE AI, TEXT-TO-IMAGE SYSTEMS AND NEGATIVE PROMPTS

Generative AI

What is Generative AI? The answer is not singular, but multiple, dispersing across realms of philosophy, art, technology and further afield; a plurality that mirrors the variety of technologies gathered under the label of artificial intelligence (AI) itself. Though generative artificial intelligence forms only a small subset of it, the conceptual ambiguity of it is also similar to its superordinate construct.

In the context of generative AI, there is a significant discrepancy between the technical and public discourse (Ronge et al., 2025). While technical discourse favours the use of the term 'generative models,' referring to its computational origins and architecture (e.g., GPT 4), a broad spectrum of semi-technical and public discourse mostly uses 'generative AI' with a system-level emphasis on the embedding of this model into a larger infrastructure (e.g., ChatGPT). In other words, generative models are machine learning architectures that use AI algorithms to draw upon patterns and relationships in training data to create novel data instances, whereas generative AI systems extend beyond these models to include the underlying infrastructure, user-facing components and their modality, as well as data processing operations such as prompts (Feuerriegel et al., 2024; Ronge et al., 2025).

One might trace the origins of generative systems to early examples such as Weizenbaum's ELIZA or Cohen's AARON

(Weizenbaum, 1966; Cohen, 1979). They can certainly be seen as formative points in the lineage but in fact, the logic of generative art precedes the technology itself. As Galanter interestingly puts it, computers did not pave the way for generative art; generative art helped pave the way for computers. That being the case, generative art is not a subset of computer art, but is rather technology-agnostic (Galanter, 2016). Think here of the Jacquard Loom, which revolutionised the generative art of weaving, and its method of punch card programming that ultimately led to the invention of computers. In this sense, generative art may be understood as any form of artistic practice based on rules or patterns, and the systems with functional autonomy that co-produces these artworks can be anything ranging from biological or chemical processes to mathematical operations or computer programs (Galanter, 2016).

That being so, a purely technical or historical approach as discussed above can hardly help us in defining what we call generative AI today. In contemporary discourse, they refer to computational systems or, more precisely, deep learning-based systems that leverage the probabilistic distribution of data to produce novel outputs in human accessible forms of text, images, video or sound. According to Ronge et al., these systems are characterised by four defining aspects, namely: (i) multimodality: the ability to operate across diverse input and output forms; (ii) interactivity: the capacity for dynamic, real time engagement; (iii) flexibility: adaptability across various tasks and domains; and (iv) productivity: the capability to autonomously generate content at a large scale (Ronge et al., 2025). This paper concerns itself with an even smaller section of generative systems that helps in text-to-image synthesis.

Text-to-Image Generative Systems

The first text-to-image systems were far from spectacular, yet, even in its crudest form, it was a radical idea (Mansimov et al., 2015) that within less than a decade we would witness the meteoric rise of the same, especially when we take into account the fidelity of the images produced or the user base of the systems in present-day contexts (Oppenlaender, 2024). This was aided by many technological and conceptual breakthroughs, including the invention of powerful graphics processing units (GPUs), the introduction of Generative Adversarial Networks (GANs), diffusion models and so forth. GANs used antagonistic parts: a generator that synthesises images and a discriminator that judges their veracity for image generation, while diffusion models operate by introducing and reversing noise in a controlled fashion (Dhariwal & Nichol, 2021; Goodfellow et al., 2014). Current generative AI systems rely predominantly on text-to-image diffusion frameworks for image creation (Rombach et al., 2021).

In this paper, we examine two widely used text-to-image generative systems: Stable Diffusion and Midjourney. Released only a few months apart in August and July 2022 respectively, these platforms have attracted large user bases, with Midjourney surpassing twenty million registered users (as of August 2025) and Stable Diffusion over ten million users (as of October 2023). When it comes to revenue, Stability AI's core model is open-source and its primary revenue streams are derived from enterprise-grade API access and its proprietary web application, DreamStudio. In contrast, Midjourney operates on a closed-source basis and generates revenue exclusively through tiered subscription plans for its user base. These models serve as critical case studies for this paper, given their scale of adoption and their role in driving discourse around generative AI.

Negative Prompts

Interaction, as noted earlier, is one of key defining features of present-day generative systems. This interaction is made possible through a number of modalities such as text, image, and sound, among others, which serve as conduits between human intention and algorithmic execution. Within this dynamic, the prompt, that is, the

text-based input assumes a greater importance especially in text-to-image generation. They function as the primary medium through which users articulate ideas, intentions or imaginaries to the generative systems.

However, this translation is never straightforward. Generating images that faithfully correspond to textual descriptions requires a fine-grained alignment between linguistic meaning and visual representation – a challenge that current generative systems struggle to meet. This makes them fall back on statistically frequent associations, omitting requested attributes or including unintended ones (Armandpour et al., 2023). This intentionality gap turns prompting into an iterative process. Users must refine and reiterate their instructions, learning to avoid ambiguity while coaxing the system toward more precise results. It is in this space of trial and error that prompt engineering has emerged as an in-demand skill. Initially coined within natural language processing for programming tasks, the term now refers more broadly to the art of crafting and refining inputs to optimise the outputs of generative AI.

Negative prompting extends this dialogue between human and machine by allowing users not only to articulate what ought to appear but also to explicitly exclude what must not (Ban et al., 2024). If the positive prompt pushes the model towards a condition, negative prompts push them away from it. Thus, rather than incorporating negations within the positive prompt like 'no trees' or 'without trees', one may specify 'trees' directly as a negative prompt. In effect, the user does not merely affirm; they carve absence with prohibitions that guide the diffusion process by subtraction. Pragmatically, it allows the user to pre-empt the burdensome labour of correcting or fine-tuning the final output through inpainting or other successive refinements, to instead guide the model away from undesired semantic trajectories from the very outset.

Despite its widespread adoption, the underlying dynamics of negative prompts remains largely opaque. As Ban et al. notes in one of the few systematic inquiries into this phenomenon (Armandpour et al., 2023; Ban et al., 2024; Du et al., 2020; Koulischer et al., 2024), two primary modalities of operation can be discerned when it comes to the realisation of negative prompts. In some cases, negative prompts exhibit a delayed effect, operating only after the positive prompt has stabilised the content of the image, as though negation were a secondary gesture, an afterthought in the unfolding of generative possibility. In other cases, negation functions through a more radical mechanism of neutralisation, wherein the undesired concept is effectively dissolved or attenuated via a mutual cancellation effect in latent space with positive prompts. Interestingly, their study also uncovers a paradoxical tendency: when applied too early in the diffusion process, negative prompts can inadvertently conjure the very detail they are meant to suppress (Ban et al., 2024). Perhaps what surfaces here is less a malfunction than a subtle reminder of how technologies carry with them their own modes of revealing.

Everything that has been discussed this far, be it on Gen AI or negative prompts, are more or less instrumentalist notions of the technology: defining them as a means to an end. In what follows, we wish to engage with a nonrepresentational account of these, where they are not treated simply as a class of objects defined by their use. For the same reason, we resist framing negative prompts merely as an affordance for erasure. Instead, we seek to engage with their underlying essence and the kind of economy they engender.

Understanding Failure

Murphy's Law states that anything that can go wrong, will go wrong. Digital technology can be the purest manifestation of this, and AI even more so. They fail spectacularly and habitually, even as we ascribe them a kind of magical infallibility (Appadurai & Alexander, 2020; Chun, 2008). Much of the contemporary scholarship underscores AI or algorithmic errors as revelatory artifacts

that make visible the ideological scaffolding of the technology itself (Ananny, 2023; Benjamin, 2019; Broussard, 2023; Noble, 2018; Wachter-Boettcher, 2017). As Ananny asserts, algorithmic errors are not found – they are made. They are products of a long history of racism, sexism, ableism and ageism in our data and technological designs (Barassi, 2024; Benjamin, 2019; Noble, 2018). This perspective identifies error(s) as a site of power: a power to dictate how systems should work, a power to declare failures, trigger fixes, and envision futures by discovering and repairing mistakes (Ananny, 2023).

Once branded a failed project that could not deliver on its lofty promises, AI now once again commands widespread enthusiasm, drawing investment and cultural fascination. And this shifting interpretation of AI as oscillating between 'winters' and 'summers' relies on a broader cultural narrative of 'rise' and 'fall,' through which societies make sense of technological transformation (Ballatore & Natale, 2023). Yet, if technological failures are to be understood as cultural narratives rather than strictly objective events, it is necessary for us to move past this dichotomy of success and failure (Appadurai & Alexander, 2020; Ballatore & Natale, 2023).

Failure is not simply an event – it is a condition, a habit. As Appadurai and Alexander note, if anything is certain, it is the certainty of failure as repetition and the repetition of failure. It is an affective economy, an infrastructural amnesia, a cycle of frustration and forgetting that sustains our engagement with these flawed systems (Ahmed, 2004; Appadurai & Alexander, 2020). Just as money (M) acquires more value through circulation to commodity (C) and back to money (M-C-M) converting it into capital, in an affective economy, affect is produced as an effect of its circulation between signs and commodities (Ahmed, 2004). In other words, affect works in a manner akin to capital through its accumulation and circulation in the social sphere. They do not reside in any specific object, sign or commodity, instead mediating the relationship between the psychic and the social, and between the individual and the collective (Ahmed, 2004; Appadurai & Alexander, 2020).

According to Appadurai and Alexander, failure functions as an affective economy by the strategic replacement of threatening ideas with less threatening ones, ensuring its continuous circulation and monetisation within digital capitalism (Appadurai & Alexander, 2020). This process influences how individuals and groups experience and interpret failure. For instance, the perpetual anxiety caused by buffering is often reframed as 'technical, soon-to-disappear nuisances' or is dismissed as user error. This displacement prevents users from recognising the precariousness, unreliability, and strategic lack of reparability of digital technologies or from largely recognising computational technology as a system of power and control. This process of 'difference and displacement' makes failure a commodity or currency where anxieties are induced and then denied or re-framed. This prevents structural, political changes by ensuring that the users remain engaged in a cycle of these repeated yet forgotten frustrations (Appadurai & Alexander, 2020).

While the sense of failure, the feeling of disappointment, regret, or remorse, is real, failure itself is not an inherent or self-evident property of systems but rather a product of judgements. These judgements are deeply shaped by arrangements of power, competence and societal expectations, raising questions about who is authorised to make such judgements and what forms they must take to appear legitimate (Appadurai & Alexander, 2020). In this sense, failure is less about the objective malfunction and more about how the malfunction is narrated, judged, and managed. Errors, glitches and breakdowns in this context become specific instances that trigger this judgement. They become frequent, often trivialised disruptions that are systematically stripped of their potential to generate critical understanding within an affective economy. Instead, they are absorbed into the system in ways that

ultimately reinforce the interests of digital capitalism.

This paper identifies negative prompts as a critical site where these error discourses become materially realised in contemporary generative systems. We examine in detail the affective-discursive construction and the imaginations of the former as a sociotechnical tool through the practices they foster and how they are entangled with the notion of failure and repair. We adopt a multimodal digital ethnography approach to achieve the same. Data was collected from these key sources: (i) user communities like subreddits (r/StableDiffusion, r/midjourney) and interviews with generative AI artists to analyse user practices and narratives surrounding negative prompts; (ii) corporate materials like API documentation, release notes etc. (from Stability AI, Midjourney) to unpack the industry rhetoric regarding errors and failures and also the framing of the said prompts as a corrective tool. Regarding Reddit posts, we purposively sampled ten posts from the subreddits r/StableDiffusion and r/Midjourney that explicitly engaged with the topic of negative prompting and which attracted significant community interaction. Although these posts formed the primary sample, the researchers had been closely following discussions in both subreddits for more than a month prior to the formal start of data collection. In addition, we conducted semi-structured interviews (via telephone) with six AI artists based in India, drawn from a larger pool of participants in an ongoing project, with each phone call lasting on average thirty minutes. They represent diverse professional backgrounds, with most working in creative industries. Their insights provided an important perspective into how such systems are taken up and reconfigured within the majority world[4]. To complement these, we also examined corporate materials, five each from Stability.ai and Midjourney, selecting only those that made explicit reference to negative prompting. Together, these materials – the interview transcripts, subreddit posts and corporate documentation – formed the empirical base for the study. The data were first subject to open coding, after which various themes such as 'progress narratives,' 'user practices,' 'prompting as labour,' 'failure as a product,' and 'affective economy of failure' are identified.

ANALYSIS

This section is structured in two parts: the first examines the corporate narratives surrounding failure and negative prompts; whereas the second turns to the user experiences of the same.

Failure as Told: Corporate Narratives of Failure and Negative Prompts

Platforms do not simply develop; they perform their development for their publics. Midjourney's announcement page[5] stages such a performance through its casual diary-like timeline. In a deliberate reframing of authority in favour of a communal environment that actively seeks feedback from users, they perform in a manner consistent with its Discord origins. However, the core narrative they put forth is no different than any other – a relentlessly positive, whiggish account of 'versions-as-evolutions' driven by a lexicon crammed with adjectives like 'improve,' 'better,' 'amazing,' 'beautiful,' and so on. These rhetorical claims are not supported by any sort of technical documentation, but through a curated spectacle of visually stunning images that functions as the primary signifier of the technology's success and capabilities. Take, for instance, Midjourney's V7 Alpha update from April 2025. It claims that the model is 'amazing,' 'smarter with text prompts,' and produces images with 'noticeably higher quality and beautiful textures,' while boasting significantly better coherence on details like hands and objects – all these claims largely unverifiable (Figure 1).

4 Shahidul Alam (2008) Majority World: Challenging the West's Rhetoric of Democracy, Amerasia Journal, 34:1, 88-98, DOI: 10.17953/amer.34.1.l3176027k4q614v5

5 https://www.midjourney.com/updates

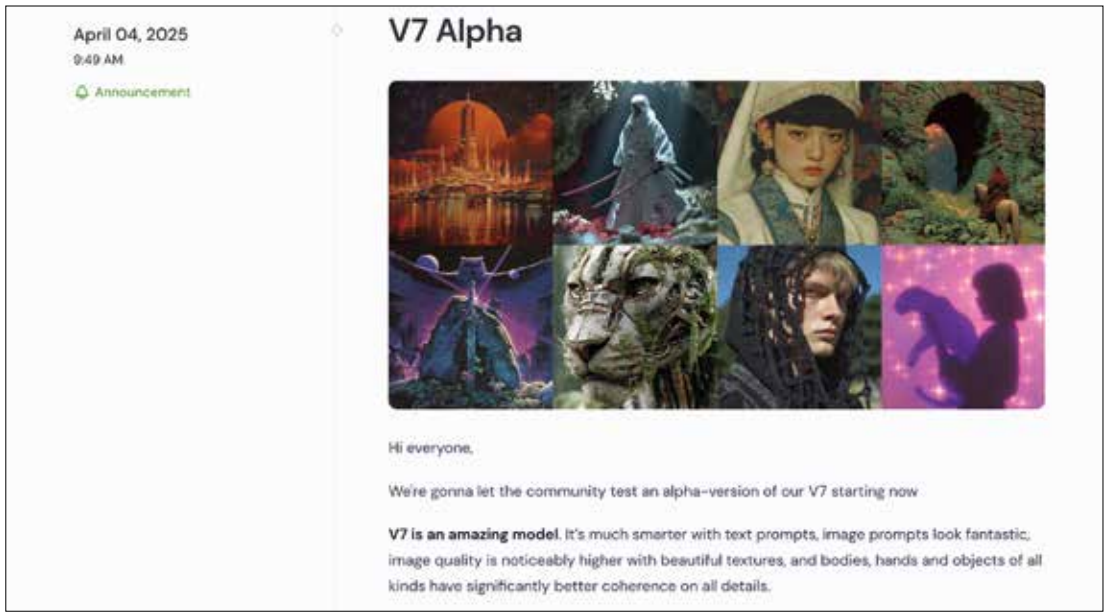

FIGURE 1. Midjourney's update on V7 Alpha

This 'superlative' development is also accompanied by a set of more-than-life visual spectacles. Such rhetoric aimed at seducing audiences to use the new model, slyly obscures systemic faults under the veneer of autonomous operations of AI. This comes from the abstracted representations that we largely glorify in the generative systems for which the real concerns such as computational power, energy consumption and human labour get submerged. The underlying reality is invoked only when they are strategically useful. For instance, it is sometimes superficially addressed to rationalise business decisions such as price increases (e.g.: "This HD mode is very expensive to run... It is ~3.2x more expensive").

That raises the following question: what, in fact, is it that they sell here? We would rather argue that it is this 'magic' that is being sold – a magic that obscures the labour on which it is built and the same that is required to sustain these systems. This framing extends to negative prompts as well. In the following part, we will aim to build on this argument in the context of negative prompts.

"It worked wonders",[6] Stability AI's hyperbolic praise grants negative prompts a kind of quasi-mythical quality, as though they were gateways to the extraordinary. For wonder is the first of all the passions, a departure from the ordinary which is not experienced or felt at all (Ahmed, 2014). Stability AI describes negative prompts as an advanced feature, a "blurb of text describing what the users do not wish to see in the output image." And in an era where intentionality, authorship and agency is contested and technical opacity is wielding power, such a framing ostensibly reaffirms control for the human in the loop. It positions negative prompts as a technical solution for the users' demand for predictable and controlled outcomes in AI-driven image synthesis. Midjourney compares this to making a 'no entry' list for 'certain' things that are undesirable in an image, ultimately appealing to the users' desire for control. This places a user who can foresee the output at an advantage – they are no longer at the mercy of these systems and can easily guide it away from happy accidents. But the question remains: who exactly holds such a position while playing with an apparatus as obscure as generative systems; and even if they do, how far does that extend? Given the paradox of the situation, what these parameters including negative prompts propose here is a false sense of mastery, an allure of control over the technology we cannot fathom.

Furthermore, this positioning rhetorically shifts the burden of precision from the apparatus to the user. Under the pretext of upskilling and convenience, these platforms foster a culture of experimentation. "Add one or more images, use multiple prompts and apply parameters to refine your results," they advise – effectively encouraging users to invest time and money in developing platform-specific competencies, while obscuring the model's benefits from them. Moreover, such a narrative asserts that the output directly mirrors the prompts' efficacy while effacing the model's own role in shaping the result. This logic extends to errors as well. With parameters like negative prompts, the users are expected to pre-empt and correct the performance errors of the system. And failing to do so is now the consequence of imprecise or ineffective prompting strategies. For instance, Stability AI, in their release note for Stable Diffusion v2.1, notes,

> "Negative prompts often eliminate unwanted details like mangled hands, too many fingers, or out-of-focus and blurry images. You can easily try negative prompts in DreamStudio right now by appending "| <negative prompt>: -1.0" to the prompt. For

6 Stability AI's update notes from December 2022 describes negative prompts as a feature that worked exceptionally well; something that 'worked wonders' in the preceding version. The sentence, verbatim, is: "Many people have noticed that negative prompts worked wonders with 2.0, and they work even better in 2.1." https://stability.ai/news/stablediffusion2-1-release7-dec-2022

> instance, appending "|disfigured, ugly:-1.0, too many fingers:-1.0" occasionally fixes the issue of generating too many fingers."

The documentation neither identifies these as errors nor investigates its underlying cause, but frames these emergent hand generation patterns in generative media as "common pitfalls of other models" or "concepts that are notoriously difficult for text-to-image image models to render," casting them as incomputable. Moreover, it is not merely the human hand that is implicated; models like Stability AI discursively construct a hierarchy that defines and labels certain features as aberrant, deviant or undesirable. What reappears here is the same technical rationality that Adorno and Horkheimer famously denounced as the rationality of domination and homogenisation. For instance, their discourse on negative prompts has come to include a broad array of features that are implicitly or explicitly coded as 'undesired,' including (but not limited to): disfigured, ugly, deformed, closed eyes, poorly drawn face, out of frame, blender, cropped, low-res, blurry, bad art, blurred, watermark, 3d render, smooth, plastic, blurry, grainy, low-resolution, anime, deep-fried, oversaturated, and the list goes on.

Implicit in these prompts are aesthetic assumptions that uphold western and ableist norms as default or desirable. Negative prompts like 'disfigured,' 'ugly,' and 'deformed' pathologise non-normative bodily or facial features, reinforcing ableist beauty ideals that equate worth with physical perfection. Similarly, labels such as 'bad art,' 'anime,' or '3D render' dismiss stylised or non-photorealistic forms, privileging western naturalism and professional aesthetics over alternative creative traditions.

What is more important to note here is the usage of 'watermark' as a negative prompt: it points to a thieving logic that these systems unabashedly endorse. This move represents a cynical attempt to sanitise the system's own derivative nature. The watermark, a signifier of the original, copyrighted source material from which these models have parasitically learned, is reframed as a mere aesthetic flaw, a noise to be subtracted through prompt whispering. And negative prompts as an affordance here become the quick fixes which can help restore the magical, glossy qualities of their outputs again. For, their 'magic' is the one of concealment – one that rests on erasure of vast data extraction practices that have become the backbone of the products they monetise.

However, a more critical reading of their positioning of negative prompts also reveals the latent framing of error (or the deviant behaviour) as the tool's default behaviour. The deviant behaviour here is assumed to be habitual, mundane and normal. They are expected, valorised, commodified, and repackaged with every update to sustain the narrative of linear progress. This cycle turns failure into a product itself. They are just another chance for these companies to revive the hype and sell the same magic solution all over again. Every update here becomes a fix on the same to maybe remain the same, as Wendy Chun might argue[7]. But marketed as a solvable problem that supposedly just takes a little AI whispering from the part of the users.

Failure as Lived: User Experiences of Failure and Negative Prompting

> "On an average, how many outputs will it take for your artworks?
> Thousands.
> Thousands? Oh my gosh.
> Yes, it goes into the thousands."

When Akhil, a Hyderabad-based AI artist, filmmaker, and author told us he literally generates thousands of images for a mere few seconds of video that he shares on Instagram, we, as researchers, were taken aback. Although a figure in the hundreds would not have come off as a major surprise for us, counts of this order, in the

7 Chun, W. H. K. (2016). Updating to remain the same: Habitual new media. The MIT Press.

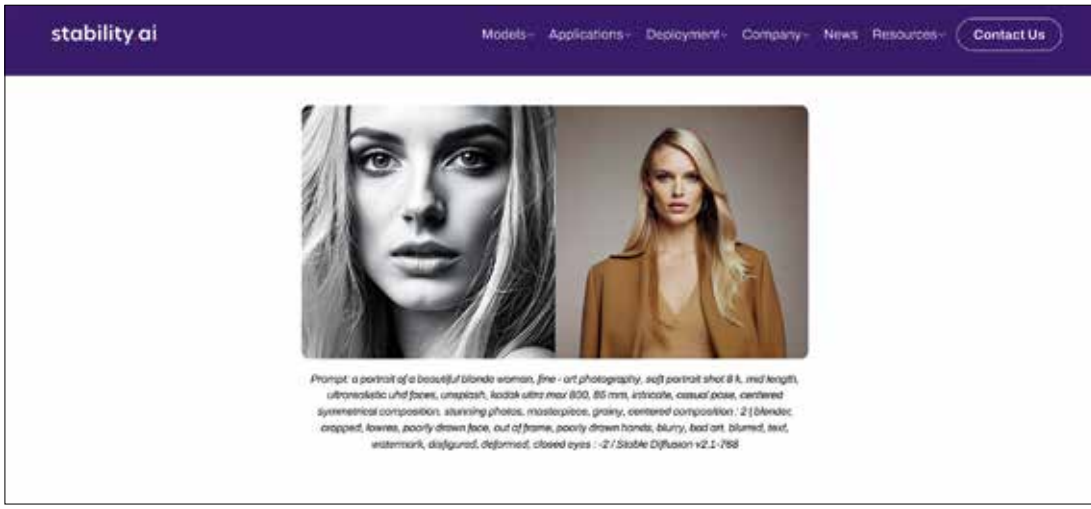

FIGURE 2. An excerpt from Stability AI's release notes for Stable Diffusion v2.1, illustrating their use of negative prompts.

thousands, were clearly beyond what we had anticipated. Perhaps, we too are guilty of a certain naivety – one that presupposes these processes to be less laborious than the traditional image-making practices we are familiar with; shaped as it is by the cultural imaginary of a sleek, efficient automation, seemingly unburdened by the drudgeries of manual practice that this industry evokes. But what unfolded in the fifty-five-minute conversation we had with him was quite the opposite; a frustrating account of the strained dynamic he had with these systems.

The initial generations, he stressed, are "very random and absolutely useless," which by his account is due to a fundamental limitation these systems have in grasping creative intent. For the same reasons, he iterates through numerous outputs until he gets what he describes as a 'realistic' output. Given the complexity of his work, it also compels him, like many others, to generate each element of an image independently – background, characters and so forth – before reworking them meticulously. Compounding this is also the fact that they are users from the majority world, who are compelled to prompt in a language that is not their mother tongue, using systems that are trained predominantly on western linguistic and cultural data. This imposes a double burden: first, navigating these systems in a non-native language; and second, the additional cognitive work of bridging cultural and contextual gaps that the model itself cannot comprehend. And even in those outputs, undesired elements creep in: nudity, figures with four eyes, or hands with ten fingers on one hand or overlapping fingers, as accounted by the artists. As Niyas, another one of our respondents, noted, "If we prompt something like a cat walking and then sitting, the output might most probably generate a tail where the leg should be – errors of that kind." This brings to foreground, again, the 'ordinary' but 'glorious' common-sense problem in AI. This is not a minor bug and never was. It is, at its core, a crisis of tacit knowledge – the unspoken, contextual expertise that humans absorb simply by moving through the world.

These errors, however silly or grave it sounds, elicited rather strong affective responses from the users. They often expressed having mixed reactions to these errors, finding them at times amusing but more often frustrating – a 'headache,' as some called it. At times, this frustration stemmed less from the errors themselves but rather from shock that a system of such 'powerful' capability could fail at "something so simple."

That makes the whole project of prompt engineering – which places the burden of performance on the human – a kind of (human) common-sense engineering. An iterative process of AI in its endless generation proposes plausible fictions from its training data and the human in the loop consistently engages in a Sisyphean task of negotiations to turn these hallucinations into a coherent reality. In a sense, it externalises the essence of GANs in flesh and silicon, with the AI systems being the generator, and the human the discriminator, with the two locked in an iterative negation and negotiation. The process has only intensified with the emergence of negative prompts, yet another product of the "frame it right, get it right" logic that these systems want to establish as intuitive judgment.

Regardless of its effectiveness, most users acknowledged that employing negative prompts was, in their view, a gamble whose outcome was just as uncertain and unpredictable as relying on positive prompts. Yet this affordance provided them with a perceived sense of control which they highly valued.

> "So I can type no nudity, don't show this, don't show that. Sometimes it obeys, sometimes it doesn't. I add things like no multiple hands, no multiple eyes. I have a very huge list of negative prompts which I add for all."

Akhil suggests framing these negative prompts as his go-to shortcut to avoid

errors. Subsequently, this 'beg crafting' or 'negative incantations,' described by one Reddit user as "strangely affective and yet ineffective," is reduced more or less to being a 'ritualistic placebo' rather than technical instruction(s). But this 'wishful thinking' is far from innocuous; under the guise of aesthetic control, these extensive lists of negative prompts that are mindlessly reproduced reinforces a hegemonic visual regime that favours certain bodies and aesthetics while terming others as aberrations. However, this temperament of the users, which allegedly and in actuality favours the system, is a part of the system, not an excuse for it (Adorno & Horkheimer, 1944).

"Bad anatomy should not be the default. If it is, then they're [the generative systems] being trained improperly," said one of the comments on a Reddit post. This showcases the users' discomfort when the digital decorum or the default is challenged. The improper here becomes the machine's misalignment with human-centric ideals of aesthetic judgement, a gap in mirroring human thought (including its biases) and its non-performance in terms of what capitalism deemed productive, efficient or acceptable. Corrective labour here becomes a burden: a burden of normativity.

Most of our respondents reported producing hundreds of iterations of the same image or video to eliminate such errors or anomalies, a process they described as both taxing and economically (as well as environmentally) burdensome.

> "They (Gen AI systems) will never understand what we are trying to say. And the worst part of this is, (even) if they don't understand, we (should) keep trying. But it costs money…And it costs me at least $150-200 per reel just for images. And video is another big headache. That cost another $150-200."

Every failed image, every wasted prompt, every instance of 'bad anatomy' becomes not just a glitch in the dream machine but also a bill to be paid. To cope, users start relying on negative prompts – not just to correct outputs, but also to save computational credits. They anticipate and negate the AI's misreadings before they even occur. And this practice is mostly driven by a posthuman doubt – an awareness of the innumerable, incalculable pathways the model could take (Amoore, 2019). This helps them trace and negate possible erroneous or undesirable trajectories before they condense into a final output. At times, this also functions as a form of pre-emptive censorship in action, where users internalise and proactively enforces platform norms. By anticipating and negating possible transgressions, such as, for instance, appending 'nudity' as a negative prompt to a benign prompt like 'showering of flowers,' users shield themselves from the financial burdens of content regeneration and the punitive reach of platform moderation.

Despite offloading most of the corrective labour onto users, this environment also trains them to look past the maintenance work they do. Users are lured into a continuous cycle of updates, releases and a promise of linear progress, such that even the most frustrated attach a cautious optimism to all their remarks. This optimism is often articulated through statements that project hope onto a hypothetical future, such as this remark: "as more Indian people engage with these tools, there might be changes – positive changes." This sentiment reflects a belief that increased demographic participation will inherently drive algorithmic improvement, a faith that persists despite experience of the system's current biases and failures. Thus, critique becomes diluted by a lingering belief in the very system's potential evolution. Everyday negative experiences are habitually counterbalanced or eased by the same forward-looking hope that sustains engagement with these systems despite persistent frustrations. This discourse also positions negative prompts as provisional tools, useful in practice 'now' but increasingly prone to redundancy as platforms evolve. As another artist observes: "I'm probably getting

erroneous outputs in only 10 to 15% of cases, while in 85 to 90%, I do not encounter these issues – particularly with Midjourney, which I primarily use." Such statements exemplify the pervasive cultivation of cruel optimism, in which users are prompted to sustain hope and continued engagement with a system whose functionality is intermittent, unpredictable and contingent on ongoing iteration. Users are thus encouraged to invest time, labour and expertise in processes that may be partially or entirely obviated by successive updates.

DISCUSSION

A toxic pinch of positivity, cruel optimism and sheer solutionism – this can define our wounded relationship with technology in general, and generative artificial intelligence systems in particular (Ahmed, 2010; Berlant, 2011; Morozov, 2014). Negative prompts as a socio-technological affordance can be the cruellest embodiment of it all – of all these utopian hopes and dystopian fears.

"Text-to-image systems have undergone drastic changes since the last time we talked [to the researchers]. They were (or are) 'faring bad' in some aspects including human anatomy, but the right training or even the right prompt can solve every problem and with that, life can be as convenient as ever" – this is the most common sentiment our respondents resonated with, a stark marker of what Morozov described as 'solutionism.' And these emotions are far from personal; it is both the product and currency of an affective economy (Ahmed, 2004; Appadurai & Alexander, 2020) that thrives on selling a linear, teleological narrative of technological development with a kind of 'solutionism' that always reaches for the answer before the questions are even fully raised (Morozov, 2014). Similar to the industry narrative that recasts AI errors as self-evident processes that can easily be optimised with negative prompts (and many more affordances that are yet to come) instead of addressing the structural issues, these tech titans of generative artificial intelligence make their 'users' a group of perennial testers – a reporter on failures whose behaviours, choices, wishes and needs are being channelled into their models (Appadurai & Alexander, 2020).

Failure here becomes a habitual and mundane part of users' everyday interactions with these systems. Neoliberalism thrives on such 'crises' and renders them ordinary. Introduction of negative prompts as an error-rectifier represents such a move – a profound admission of failures made ordinary. It has produced a class of super-empowered subjects called upon to intervene decisively and steer situations toward resolution, like the class of artists being taught to tackle the systemic failures by 'engineering' (negative) prompts. With these crises becoming ordinary, it forecloses the door for real change, producing a present that is nothing but an impasse – an affectively charged cul-de-sac (Berlant, 2011; Chun, 2016).

Meanwhile, failure also becomes a commodity, as a strategic means of valorisation, one that operates simultaneously as a value-generating product for its proprietors. It becomes intimately linked to innovation, growth and profitability (Appadurai & Alexander, 2020). Each update of these generative systems strategically capitalises on this narrative, effectively monetising and leveraging the very failures of previous iterations. As Wendy Chun suggestively frames it, for new (habitual) media, they live and die by the update: the end of the update, the end of the object. For them, 'to be' is to be updated. Habit + Crisis = Update; or put differently, program (X) + exception (X) = program (X + 1) (Chun, 2016). Generative systems are no exception: they thrive on the same calculus and hence the relentless procession of versions: V1, V2, V3, V4, V5, V5.1, V5.2, V6, V6.1, V7 – each promising novelty and upgrade, each little more than failures repackaged as evolution.

This update culture continuously disrupts stability, fostering new habits of dependency , perpetuating a cycle of perceived urgency and necessary recalibration (Chun, 2016). It also helps frame and reframe the negative experiences of dissat-

isfaction and frustration on the part of the users in order to maintain the continuous repetition of failure (Appadurai & Alexander, 2020). Their constant displacement of ideas and the strategic production of forgetfulness that these systems endorse often displaces the user's anxiety with a kind of cruel optimism, which was evident with the user's outlook towards AI bias, performance errors, negative prompts as an affordance or other aspects of the system. This anticipation is coupled with knowing that disappointment of the prosumer is what drives the systems' ephemeral yet enduring nature (Chun, 2016). It is this forgetfulness, the concealment of labour, both past and present, that is the magic that these systems try to legitimise.

As Lotringer and Virilio put it, to invent the train is to invent the train crash, to invent the plane is to invent the plane crash (Lotringer & Virilio, 2005); and by extension, to invent generative AI is to invent its ever-present failures – the malfunctions, the errors, the unforeseen breakdowns coded into its very operation. It is never, and will never be, 'ready' to hand (Graham & Thrift, 2007). It demands its maintainers. In the liminal space between breakdown and restoration – between the visible (broken) tool and the inconspicuous one – exist repair and maintenance. We argue in this context that AI artists are not merely users of these generative systems but also repairers, the maintainers of it. It is largely their tacit knowledge that maintains these systems – its glossiness and magic. However, often concealed by the age-old narrative that attributes all successes to technology while displacing responsibility of failure to the user, the consumer and an affective economy that valorises and commodifies failure. But what is it that is being maintained here? Is it the system or the thing itself, or the negotiated order that surrounds it, or some 'larger' entity? (Graham & Thrift, 2007). That is still a difficult question to answer.

CONCLUSION

Failure is never a one-time event. It is a commodity exchanged, monetised and valorised, perhaps most strikingly in the context of today's generative AI systems which are always oscillating in a circuit of human expectation and machinic creativity. This dynamic fosters an affective economy built precisely on failure. They function by making errors ordinary and transferring the burden of their elimination onto the users – most conspicuously enacted through the feature of negative prompts in visual generative systems. Here, the users are made to perform the dual role of consumer and unpaid trainer. They become a class of persistent testers who report, repair and maintain the system whose labour of making sense, of bridging gaps, of repairing breakdowns is continually overshadowed by the enchantments we attach to machines and are rendered invisible beneath the corporate fable of linear, steady and unstoppable progress.

We began this paper by invoking the magic of generative AI, a vision of effortless creation that captivates both cultural imaginations and industry rhetorics. We want to end it with a reminder that, like every magic, it is also sustained by labour and effort that is often obscured. As one of our respondents remarked, "It cannot magically generate [something] out of its hat. You need to train it, you have to teach it."

ACKNOWLEDGMENTS

We are greatly indebted to all the AI artists who generously shared their time and insights with us. This study would not have been possible without their contributions. We also thank the reviewers of this paper for their valuable feedback and suggestions.

REFERENCES

Adorno, T., & Horkheimer, M. (1944). The Culture Industry: Enlightenment as Mass Deception.

Ahmed, S. (2004). Affective Economies. Social Text, 22(2), 117–139. https://doi.org/10.1215/01642472-22-2_79-117.

Ahmed, S. (2010). The Promise of Happiness. Duke University Press. https://doi.org/10.2307/j.ctv125jkj2.

Amoore, L. (2019). Doubt and the Algorithm: On the Partial Accounts of Machine Learning. Theory, Culture & Society, 36(6), 147–169. https://doi.org/10.1177/0263276419851846.

Amoore, L. (2020). Cloud ethics: Algorithms and the attributes of ourselves and others. Duke University Press.

Ananny, M. (2023). Making Mistakes: Constructing Algorithmic Errors to Understand Sociotechnical Power. Osiris, 38, 223–241. https://doi.org/10.1086/725146.

Appadurai, A., & Alexander, N. (2020). Failure. Polity press.

Armandpour, M., Sadeghian, A., Zheng, H., Sadeghian, A., & Zhou, M. (2023). Re-imagine the Negative Prompt Algorithm: Transform 2D Diffusion into 3D, alleviate Janus problem and Beyond (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2304.04968.

Ballatore, A., & Natale, S. (2023). Technological failures, controversies and the myth of AI. In S. Lindgren (Ed.), Handbook of Critical Studies of Artificial Intelligence (pp. 237–244). Edward Elgar Publishing. https://doi.org/10.4337/9781803928562.00026.

Ban, Y., Wang, R., Zhou, T., Cheng, M., Gong, B., & Hsieh, C.-J. (2024). Understanding the Impact of Negative Prompts: When and How Do They Take Effect? (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2406.02965.

Barassi, V. (2024). Toward a Theory of AI Errors: Making Sense of Hallucinations, Catastrophic Failures, and the Fallacy of Generative AI. Harvard Data Science Review, Special Issue 5. https://doi.org/10.1162/99608f92.ad8ebbd4.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the New Jim Code. Polity.

Berlant, L. (2011). Cruel Optimism. Duke University Press. https://doi.org/10.2307/j.ctv1220p4w.

Broussard, M. (2023). More than a glitch: Confronting race, gender, and ability bias in tech. The MIT Press.

Chesher, C., & Albarrán-Torres, C. (2023). The emergence of autolography: The 'magical' invocation of images from text through AI. Media International Australia, 189(1), 57–73. https://doi.org/10.1177/1329878X231193252

Chun, W. H. K. (2008). On "Sourcery," or Code as Fetish. Configurations, 16(3), 299–324. https://doi.org/10.1353/con.0.0064.

Chun, W. H. K. (2016). Updating to remain the same: Habitual new media. The MIT Press.

Cohen, H. (1979). What is an image. International Joint Conference on Artificial Intelligence. https://api.semanticscholar.org/CorpusID:62789816.

Dhariwal, P., & Nichol, A. (2021). Diffusion Models Beat GANs on Image Synthesis (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2105.05233.

Du, Y., Li, S., & Mordatch, I. (2020). Compositional Visual Generation and Inference with Energy Based Models (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2004.06030.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. Business & Information Systems Engineering, 66(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7.

Galanter, P. (2016). Generative Art Theory. In C. Paul (Ed.), A Companion to Digital Art (1st ed., pp. 146–180). Wiley. https://doi.org/10.1002/9781118475249.ch5.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1406.2661.

Graham, S., & Thrift, N. (2007). Out of Order: Understanding Repair and Maintenance. Theory, Culture & Society, 24(3), 1–25. https://doi.org/10.1177/0263276407075954.

Koulischer, F., Deleu, J., Raya, G., Demeester, T., & Ambrogioni, L. (2024). Dynamic Negative Guidance of Diffusion Models (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2410.14398.

Larsson, S., & Viktorelius, M. (2024). Reducing the contingency of the world: Magic, oracles, and machine-learning technology. AI & SOCIETY, 39(1), 183–193. https://doi.org/10.1007/s00146-022-01394-2.

Lotringer, S., & Virilio, P. (2005). The accident of art. Semiotexte.

Mansimov, E., Parisotto, E., Ba, J. L., & Salakhutdinov, R. (2015). Generating Images from Captions with Attention (Version 2). arXiv. https://doi.org/10.48550/ARXIV.1511.02793.

Morozov, E. (2014). To save everything, click here: The folly of technological solutionism (Paperback 1. publ). PublicAffairs.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York university press.

Oppenlaender, J. (2024). The Cultivated Practices of Text-to-Image Generation. In R. Rousi, C. Von Koskull, & V. Roto (Eds.), Humane Autonomous Technology (pp. 325–349). Springer International Publishing. https://doi.org/10.1007/978-3-031-66528-8_14.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-Resolution Image Synthesis with Latent Diffusion Models (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2112.10752.

Ronge, R., Maier, M., & Rathgeber, B. (2025). Towards a Definition of Generative Artificial Intelligence. Philosophy & Technology, 38(1), 31. https://doi.org/10.1007/s13347-025-00863-y.

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. American Psychologist, 26(2), 145–159. https://doi.org/10.1037/h0030806.

Wachter-Boettcher, S. (2017). Technically wrong: Sexist apps, biased algorithms, and other threats of toxic tech (First edition). W.W. Norton & Company.

Weizenbaum, J. (1966). ELIZA – a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36–45. https://doi.org/10.1145/365153.365168.