

Generative AI in Film: CGI fatigue as a harbinger of another AI winter?

Written by Samuel Flückiger

As a filmmaker, avid moviegoer and consumer of various film industry media who grew up in the 80s and 90s, I'm lucky to have observed and experienced one of the most exciting and disruptive technological advances in cinema since the advent of sound:

The emergence of computer-generated imagery (CGI) has fundamentally changed film production, offering a vast set of tools that enables filmmakers - independent and studio-employed alike - to tell stories, which were labeled unfilmable for most of cinema's history. Now, a second disruption looms with a potential technological advance *within CGI*, which some proponents believe will change the history of cinema yet again: the incorporation of *machine learning-based image generators* into CGI, to either replace the process of creating entire frames - what I will call the *generative full-frame approach* - or to at least replace legacy visual effects techniques - what I will call the *assistive tool approach*:

- (1) Generative full-frame approach: The complete replacement of the camera as the cinematic engine with machine learning-based image generators. Examples: MidJourney, Stable Diffusion, Sora, etc.
- (2) Assistive tool approach: The complete replacement of legacy processes within post-production with machine learning-based generative tools. Examples: Generative fill in After Effects, machine learning-based rotoscoping in Davinci Resolve, etc.

I would like to honor the arguments of Jijang et al in their text «AI Art and its impact on artists» by not simply referring to «AI» in this essay, and rather use the more specific term *machine learning-based image generators*, as it struck me as vital to demystify the term AI and somewhat counter its prevalent and ongoing anthropomorphisation; especially in the context of cinema; since there is a lot of fear within the filmmaking community rooted in the lack of understanding how these machine learning-based image generators work; what their limitations are and how much of a threat they actually represent to the creative process and the livelihood of filmmakers, designers and technicians. Suchman's text «The Uncontroversial 'Thingness' of AI» further underlined this argument and also had a significant impact on me as a reader.

One of the films that spellbound me in my youth and made me want to become a filmmaker was the Wachowski siblings' 1999 science fiction film «The Matrix» - which conveniently serves as a great example for this essay, since it tells a story of a rebellion against an all-encompassing AI that enslaves humans, turning them into organic battery storage.

«The Matrix» not only offers an all-too-relevant metaphor that works well in today's sociopolitical context of «AI-fueled» social media and LLMs allegedly numbing people into social apathy and declining cultural participation. It also serves as an interesting example of a film that still relied on combining CGI with in-camera, practical effects, while its sequel; «The Matrix: Reloaded» (2003) became one of the first films to use a fully virtual camera for certain scenes; most famously in the «burly brawl» scene, which transitions to full-frame CGI when the hero of the film, Neo (played by Keanu Reeves) fights hundreds of incoming copies of his nemesis, Agent Smith (played by Hugo Weaving).

«The Matrix» works well for me here to reflect on what has become a highly debated topic in the filmmaking community: the so-called *CGI fatigue*; which - as I will argue later on - might be a harbinger of an incoming (or already spreading) *AI fatigue* plaguing audiences' reception in media overall but in cinema particularly.

«The Matrix: Reloaded» was my first encounter with fully CGI-generated footage, and I remember vividly being thrown off and disappointed by the lack of photorealism of the «burly brawl» scene: textures, especially those of human skin and textiles, as well as the scene's physics looked more like those from a computer game, rather than what I was used to from the first «Matrix» film; where real actors were captured using real cameras even for the most complex, time-warping slow-motion shots (also known as «bullet time»).

The subdued critical response of the «Matrix» sequels reinforced my disappointment. And while people still showed up at the box office, I argue that the reason for the perceived artistic shortcomings of these films is partly rooted in the use of full-frame CGI:

I claim that within these fully CGI-generated shots, *nothing is at stake*. And by that I also mean: *Nothing non-diegetic is at stake*.

So it's important to note that I don't suspect the actual lack of physical realism or detail in textures per se to be the reason why I was thrown off by these full-frame CGI shots. Rather, I believe that the absence of *a deeper cultural value* is the root cause for these first symptoms of CGI fatigue.

And before I get into what I mean by that; I need to state that my claim of CGI fatigue is an entirely personal and hence subjective assumption; and this trend would have to be scientifically proven first. But I think there are some signs that this observation has substance: one could argue, for example, that the decline of the «Marvel» universe is directly linked to the idea of CGI fatigue, or that further proof lies in the success of filmmakers who almost religiously decline working with full-frame (or even assistive tool) CGI such as Christopher Nolan or Paul Thomas Anderson.

So what do I mean by describing *stakes* - and especially *non-diegetic stakes* - as the key to successful image creation and filmmaking as a whole?

According to the Merriam-Webster Dictionary, a «stake» is something that «is put up to be won or lost in a contest or endeavor», which - and this is my hypothesis - in filmmaking and the subsequent reception of the outcome of filmmaking; a shot, scene or a film; translates into two components:

- (1) Diegetic stakes; e.g. «Neo must free Morpheus from the agents before they torture him to death, otherwise the rebellion against the machines is doomed».
- (2) Non-diegetic stakes; e.g. «actors Keanu Reeves and Hugo Weaving had to undergo grueling martial arts training before the shoot. Otherwise, their fighting scenes would look dull and laughable».

While the first type of stakes represents the standard prerequisite for a successful scene or film - if nothing is at stake for the characters, then it will not evoke an emotional response from the audience - the second form of stakes is as crucial, especially but not only for big blockbuster films: Non-diegetic stakes make the cinematic experience exciting, even if diegetic stakes are absent: The spectacle of the production itself; directors, actors and designers putting themselves into dangerous situations and/or risking their health, lives and/or sanity; while investing their artistry into the cinematic product. I claim that computer-generated imagery inherently lacks these stakes.

Why? CGI represents an avoidance of creative risk taking and the willingness to put anything but time and money at stake. And it is my suspicion that this becomes ultimately tiresome for audiences - hence the term *fatigue*.

Now what does all of this have to do with machine learning-based image generators?

I claim that the arguments above outlining *CGI fatigue* fully apply and directly translate into what will eventually become *AI fatigue*:

Even if machine learning-based image generators reached visual perfection in full-frame generation and all their current challenges of body integrity, physics, training bias, consistency across scenes, etc., would be completely overcome; the absence of non-diegetic stakes will always negatively affect the generated shots (and ultimately scenes and entire films) in such a significant way, that audiences will eventually grow tired of them.

Combining this idea with the growing fear of a bursting economic bubble of massively overvalued AI-themed companies, I argue that AI fatigue in cinema could make a small but significant contribution to the arrival of another *AI winter*; a historical concept that is outlined

in the text «Why AI is harder than we think» by Melanie Mitchell, which describes a repeating cycle in AI history in which it loses prominence and significance due to unfulfilled technological promises and collapsing funding.

Of course it is hard to argue against the notion that machine learning-based image generators have become so successful due to the availability of big data and the enormous computing power of today's GPUs, that they will always remain part of the cultural ecosphere and thus filmmaking. But I strongly believe that the *generative full-frame approach* will eventually lose significance due to the argued lack of non-diegetic stakes; and that it will more likely be the *assistive tool approach* that will prevail when it comes to the future of machine learning-based image generators.

And since through the *assistive tool approach* machine learning-based image generation is deployed in a much more curated way - e.g. designers still design when using machine learning-powered tools, instead of being wholly replaced by image generators themselves -, its inclusion in the cultural context of various art forms will be a more organic, less disruptive and much more accepted development - by audiences and creatives alike. The *assistive tool approach* still leaves enough room for humans to risk something; to put something at stake; which ultimately imbues the outcome of the creative process with more cultural value than any outcome of the generative full-frame approach ever could.

So while I suspect that the return of an AI winter mainly will depend on the continued success and economic stability of companies selling machine learning-based generators of all kinds within broader aspects of society (mainly the economy); the audience has a role to play, and I argue that if the audience grows tired of machine learning-based content due to the absence of culturally significant, non-diegetic stakes, AI in the form of full-frame image generators is destined to face a winter - how warm or harsh remains to be seen.

To conclude, I have not yet decided on what to do in my CAS project, but a profound interest has been sparked to further explore how resilient the art form of cinema is in the face of machine learning-based image generators.