Chances and Limits of Immersive Environments for Anti-Discrimination and (Historical-)Political Education

Deborah Schnabel

The digitalization of the public sphere puts political educators under considerable pressure. With every new technology and app, a new forum emerges that can become a problem for anti-democratic agitation but also offers the opportunity to make democracy education better for more people. Racism, anti-Semitism, and other threats to democracy are present in all areas of life, and they are quickly conquering the new digital worlds. With every new space, democracy is also renegotiated, and political education is called upon to intervene in this negotiation.

Immersive environments (IE) offer potential to reach more people at a lower threshold and with a different degree of involvement. There is great interest from the entire field of political education to become effective in these new worlds. It may even be possible to preserve important testimonies and witnesses of history by recreating them in virtual realities. Therefore, it is primarily historical-political education, memorial sites, and museums that are launching projects despite the insufficiently explored possibilities of immersive environments.

Especially memorials feel the pressure to make places of remembrance alive, and of course, especially in the field of Shoah Education: the last eyewitnesses are leaving us. First-hand accounts will no longer be possible. Many actors in political education see Immersive User as a possibility to reconstruct the authenticity of a contemporary witness encounter. For example, there is the Anne Frank House in Amsterdam, whose digital VR exhibition allows a tour of the family's hiding place. There is also the app "Inside Auschwitz", developed by WDR, which combines the accounts of three eyewitnesses with a virtual visit to the memorial (Planet Schule, 2022). Additionally, there is the project "Dimensions in Testimony" by the USC Shoah Foundation (USC SF), which lets eyewitnesses speak to us as AI-based holograms (Dimensions in Testimony, n.d.). There are also the first projects in which students "travel back" in the metaverse to key moments in Black history (Education in the Metaverse, 2023). What all actors in the field have in common is that they perceive the new possibilities

of IE as very important, sometimes invest a lot in the field, have high expectations of the possibilities, but at the same time deal with them rather restrictively. For example, the VR Anne Frank House has no protagonists at all, but instead presents the hiding place furnished, in contrast to the physical exhibition. Other applications only allow certain characters, who usually have a bystander role. Behind this lies the still unanswered question of how to unite the possibilities of new technology with the principles of political education. It will depend on the quality of the answers whether IE will become suitable for widespread use in democracy education and history teaching.

No matter how it is approached, IE will come up against limits – technological ones (which I will address later), but often also educational-theoretical ones: Is "artificial authenticity" even possible? Does it make pedagogical sense? Or is an immediacy that cannot be reconstructed being fetishized here?

How Realistic Can Remembering Be?

The limits of immersion are often discussed in such applications, a debate that is older than IE. In memorial education, long before the Internet, there was a debate about how deep the experience of the visitors should go. The main issue was whether such an experience should be accompanied or unaccompanied. In German-speaking contexts, the so-called Beutelsbach Consensus was used as a guideline, which emphasizes the prohibition of overpowering, the requirement of controversy, and student orientation. It is difficult to apply Beutelsbach to IE, as such environments usually overwhelm from their very game logic. In the USA, however, the focus is often on "empathy education", the development of compassion and understanding, often with methods that are quite "overwhelming" (Besand et al., 2019). The problem of "artificial authenticity", which can lead to distancing or empathy, is obvious: history can potentially be distorted. A purely imagined image of the past emerges that is neither congruent with historical knowledge nor with the victims' world of experience.

Apart from the inherent pedagogical risks of the technology, the controversy also goes to the content. The biggest point of contention is the possibility for non-victims to play victims of racism or anti-Semitism in IE. What some may view as an opportunity for empathy and a change of perspective can be seen by others as insensitive, cultural appropriation,

false alignment of perpetrators and victims, violation of the prohibition of overwhelming and a potential trigger for (transgenerational) trauma. Complex feature compositions are truncated and cannot be explored in depth, especially when, for example, a typical VR session usually does not exceed one hour. The socio-political effect on the dominant society must also not be underestimated: Memory-political problematics and exoneration narratives can be reinforced if the descendants of the perpetrators can feel even more strongly as part of the persecuted through IE.

The balance of distance and empathy that political education has to maintain in IE could be understood as a kind of variant of the "uncanny valley" effect – at a certain level of reality of the simulation it becomes strange, "creepy". The mentioned example of the WDR app "Inside Auschwitz" illustrates how difficult this balancing act is. On the one hand, it deliberately tries to eliminate the distance: "The users should consciously not take a distanced position but explore the memorial site as if they were visiting it on foot." (Nägel & Stegmaier, 2019). On the other hand, the app's producers warn against too much closeness and even recommend not using VR glasses in a school context, as the lack of distance can disturb students.[1]

It appears that applications are better accepted when they use alienation effects, creating a mixed space that is not real, but not completely fictional either. It is not about identification, but rather sympathy. This seems to apply to the medIEm as a whole: while VR/AR are still niche products, MMORPGs[2], for example, are functioning excellently as a mass medIEm. Here, it often appears to be the artificiality of the environment that makes it possible to deal with it confidently. There is never any doubt that one is not in reality; an avatar that does not resemble the user enables greater identification than the clinical perfection of VR/AR, which often does not enable the user to experience their body at all, as when looking down with VR goggles, one does not see their body, but the floor.

It seems almost impossible to consider every potential user and every possible multiple concern. Can't we operate similarly to our everyday experience – with limited knowledge, but with room to learn? Our daily

1 "The stories told by the witnesses and the images from the camp can be challenging for sensitive students or young people with their own experience of war and flight. It can help to watch the videos with a little more distance (for example, over the shoulder of a classmate holding the tablet). In addition, it is advisable not to use headphones or VR." (Planet Schule, 2022).

2 Massively Multiplayer Online Role-Playing Games, online video game worlds with many thousands of users.

behaviour is also based on limited knowledge, yet we are getting very far here as well.

Educational science findings on the pedagogical impact of IE are still manageable: mixed results were found for "Anne Frank House VR" compared to desktop learning – while identification with those involved was stronger, at the same time less content was absorbed (Mulders, 2023). Learning dynamics are complex and require more intensive research.

Metaverse and Beyond

Certainly, the best-known IE application is Mark Zuckerberg's "Horizon Worlds", widely known simply as "Metaverse"[3] – known both in terms of ambition and for negative headlines. What is less important for our field is whether "the Metaverse," i.e., Horizon Worlds has been successful, but it is uncertain whether a similar platform for IE applications could be successful. Civic education must prepare for the possibility that such a platform could become a mass medIEm. Comparisons can be drawn with the failed "Second Life" platform.

The same questions that apply to social media also apply to the metaverse: Who leads the discourse? Who dominates? Do discriminated people feel safe in the metaverse? The higher degree of reality of IE puts the classic problems of social media into sharper focus. There is still little research on whether IE has an even greater potential for radicalization than social media already does. Heuristically, it is probably safe to assume a similar threat level as social media. Furthermore, adaptive and reactive content in IE will increasingly be produced by AIs, as human content development

3 The term "metaverse" in the narrower sense is theoretically meant to encompass all immersive environments and their interconnection – in this sense, "the metaverse" does not yet exist, but individual IEs Horizon Worlds, Minecraft, etc. do. ChatGPT offers the following definition of the metaverse (prompt: write a scientific definition of the metaverse): „The Metaverse is a comprehensive, large-scale virtual environment that integrates multiple interconnected digital spaces, facilitating seamless interactions, communication, and collaboration among users, who are represented by digital representations known as avatars. This virtual framework merges aspects of virtual reality, augmented reality, and interconnected networks to create an immersive and persistent ecosystem for users, with wide-ranging applications spanning entertainment, social interaction, education, commerce, and various other domains. By transcending physical limitations and geographical boundaries, the Metaverse fosters novel opportunities for creative expression, innovation, and interaction in the digital realm."

cannot keep up with the demand for these demanding environments – this is also the view of a contribution by the World Economic Forum, for example (AI Is Shaping the Metaverse, 2023). In this scenario, it is conceivable that AI language models could create personalized worlds that essentially affirm users' attitudes, reinforce them in problematic attitudes, force ideological cocooning effects, and drive them into radicalization tunnels. The already low inhibition threshold in the metaverse would do the rest.[4] These structural aspects of AI racism could have an exponential effect in immersive environments: Racist echo chambers in virtual worlds would inevitably leak outward, shaping society – and in turn serving as a source of AI content generators.

The likelihood of this scenario also depends on the people who develop the metaverse. They would need to be diversity- and discrimination-sensitive from recruitment through beta testing in all areas of work. Experience has shown that this cannot be achieved through decree, but through the active participation of the people involved. The question of who uses the metaverse and why has not yet been answered. What inhibitions might exist for certain parts of society to use the Metaverse? Why might this be the case? Do corporations like Meta actively counteract such tendencies? What is being done to provide representation and protection? How do Meta and Co. prevent right-wing infiltration, as can also be observed on other digital platforms that operate more in the niche?

Political Education in the Metaverse

Apart from these questions, this task arises in our field of action: How should civic education respond to these and similar questions? In our opinion, the question should not be whether political education should now "invest in the metaverse." Such a generalist approach is doomed to fail. However, it should be experimented with – the metaverse should be seen as software in which concrete, limited settings are tried out. Thus, the maxim should not be "We're putting our whole memorial into the Metaverse," but rather, "This specific exhibition can (also) be visited in the Metaverse."

4 cf. only Yinka Bokinni: A barrage of assault, racism and rape jokes: my nightmare trip into the metaverse. (25.04.22.) URL: https://www.theguardian.com/tv-and-radio/2022/apr/25/a-barrage-of-assault-racism-and-jokes-my-nightmare-trip-into-the-metaverse (12.09.23)

The potential of such steps can already be observed today, for example in Minecraft. For example, a library of black literature was saved not only to Meta's "Horizon Worlds" but also to a Minecraft server. One explanation for Minecraft's success could be found in the "uncanny valley," as it does not try to imitate reality. There is currently a project underway to transfer the Yad Vashem memorial to Minecraft (Yad Vashem Comes to Minecraft, 2023), as well as private projects that strive to reflect the Holocaust on the platform (Jcirque25, 2020). However, this should not hide the fact that Minecraft is also used by the opposite side: users often point out servers on which Nazis glorify Nazi atrocities.

Excursus: AI, Discrimination and Justice

The increasing importance of AI in various environments is platform-independent and requires its own analysis. The language used to discuss "artificial intelligence" promises a revolution in all aspects of life, from the world of work and macroeconomics to literature and art. At the same time, a promise of justice based on procedural rationality and the absolute neutrality of the algorithm is being made. Exclusions resulting from misanthropic attitudes can thus be avoided automatically – the assessment is made on the basis of an alleged objectivity that cannot be achieved by humans.

In fact, the publicly known language models, first and foremost ChatGPT, do not seem to be suitable for reproducing discriminatory language – ChatGPT in particular stands out here for its robust ethical framework. ChatGPT strongly opposes racism and discrimination. All attempts to use the software to produce racist texts are supposed to fail theoretically already in the approach (although this can be circumvented with the appropriate Prompts). It should not be forgotten that there are projects that specifically train language models to combat misanthropy, such as the international research project "Decoding Antisemitism"[5].

At the same time, there are increasing reports that structural racism is being reinforced by intelligent image recognition software. The algorithm bases its decisions on patterns it finds in society and evaluates non-white people differently than white people. This means that racism is avoided on a linguistic level, while it is further perpetuated, automated, and industrialized on an operational level, usually without transparency for those affect-

5 https://decoding-antisemitism.eu/ (12.09.23).

ed. If AI is used to make decisions about credit and housing allocation, job applications, and social forecasts in the future, the problems are foreseeable. Unlike human decisions, the apparent objectivity of the machine still makes it immune to criticism.

The framework conditions of the AI industry also need to be critically examined – for example, the underpaid Kenyan workers whose performance was essential for the success of ChatGPT (Perigo, 2023). On an economic level, AI reproduces capitalism instead of critically questioning it, and this is done along racial distinctions. Who owns the systems? Who is allowed to market their services, and who benefits from the increased productivity? Who owns AI art, whose creation is inspired by the names of famous artists? Who can work in this new AI world, whose job may be replaced by automation, and whose life may be managed by AI? All indications suggest that those affected by racism will not receive positive answers to these questions, and that the negative effects of the revolution will disproportionately affect them.

Furthermore, the implications for antisemitism should not be underestimated. Already, the portrayal of Jewish people is subject to a historical bias – image-creation programs draw primarily on historical representations and images of ultra-Orthodox groups. As a result, Jewish people, who today participate in modern life as a matter of course, are absent from the collective imagination.

Last but not least, Artificial Intelligence (AI) systems can already be used to generate an endless stream of fake news, deep fakes, and conspiracy theory content – with a plausibility and apparent authenticity that current conspiracy ideologues can only dream of. Likely, the effects of this will be felt particularly strongly by Jewish people.

All these questions arise with increased urgency when AI is not just a nice toy within Information Units (IEs) but is instead constitutive of them. How will educational IEs change when they no longer present content that is prefabricated by humans, but instead AI dynamically generates such content? What will happen when AI-based avatars target users and radicalize or intimidate them with personalized propaganda? Where is the control? And how can automated patterns of reproduction of anti-human attitudes be interrupted?

This requires asking questions that naturally point to the need for further exploration. Beyond the field of IE in civic education, namely: what conditions for success do we need for a discrimination-sensitive AI ethics? How can AI be taught to recognize power asymmetries? Here, limits become

apparent that are already inherent in the current way of working of AI – it is fundamentally reproducing.

A Counter-Model: IE as Lived Utopia

Perhaps it is not expedient to reconstruct stories of persecution and experiences of discrimination for political education in IE. Perhaps it would make sense to simulate worlds in IE in which a certain utopian hope has become real. Instead of putting learners into a pedagogical capsule with certain preconceived content and learning goals, they could independently explore worlds that have already been liberated. The learning effect would arise precisely from the absence of a context of violence that is now commonplace.

An example of this is the everyday simulation game "The Sims" (released in 2000) and its role in promoting homosexual/queer emancipation. In this PC game, queer relationships could be experienced for the first time in a mainstream title without it being a constant theme. Many queer people today report having experienced queer representation and empowerment for the first time simply because of the possibility that was not imposed or used as a teaching tool but was explored and motivated intrinsically.

Perhaps political education for IE should take a similar approach: provide a concrete utopia that can be explored, with possibilities users can experiment freely – and in the process reduce fears and prejudices. The virtual world should be a model for reality, not the other way around.

This approach would possibly have the advantage of avoiding the numerous pitfalls mentioned above that threaten political education in IE from the outset. The boundaries of the user experience would also have to be drawn less tightly; there would be less need for pedagogical guidelines, and less top-down control overall.

References

AI Is Shaping the Metaverse – But How? Industry Experts Explain. (2023, May 9). World Economic Forum. https://www.weforum.org/agenda/2023/05/generative-ai-and-how-can-it-shape-the-metaverse-industry-experts-explain/

Besand, A., Overwien, B., & Zorn, P. (Eds.). (2019). *Politische Bildung mit Gefühl*. Bundeszentrale für politische Bildung.

Dimensions in Testimony. (n.d.). Speak with survivors and other witnesses to the Holocaust and other genocides through their interactive biographies. USC Shoah Foundation. https://sfi.usc.edu/dit

Education in the Metaverse Takes Students Back in History. (2023, December 1). WBUR. https://www.wbur.org/hereandnow/2022/12/01/black-history-metaverse

Jcirque25. (2020). Using Minecraft as a medIEm to reflect on the Holocaust. *Reddit*. https://www.reddit.com/r/Minecraft/comments/htlst6/using_minecraft_as_a_medIEm_to_reflect_on_the

Mulders, M. (2023, March 6). Learning about victims of the Holocaust in virtual reality: The main, mediating, and moderating effects of technology, instructional method, flow, presence, and prior knowledge. *MDPI*. https://www.mdpi.com/2414-4088/7/3/28

Nägel, V. L., & Stegmaier, S. (2019, October 8). AR und VR in der historisch-politischen Bildung zum Nationalsozialismus und Holocaust – (Interaktives) Lernen oder emotionale Überwältigung? *Bundeszentrale für politische Bildung*. https://www.bpb.de/lernen/digitale-bildung/werkstatt/298168/ar-und-vr-in-der-historisch-politischen-bildung-zum-nationalsozialismus-und-holocaust-interaktives-lernen-oder-emotionale-ueberwaeltigung/

Planet Schule. (2022, December 10). Inside Auschwitz | Erzählungen Überlebender und 360 Grad-Videos. *Planet Schule*. https://www1.wdr.de/schule/digital/unterrichtsmaterial/dreisechzig-inside-auschwitz-100.html

Perrigo, B. (2023, January 18). OpenAI used Kenyan workers on less than $2 per hour to make ChatGPT less toxic. *TIME*. https://time.com/6247678/openai-chatgpt-kenya-workers/

Yad Vashem Comes to Minecraft for International Holocaust Day. (2023, January 26). i24NEWS. https://www.i24news.tv/en/news/israel/technology-science/1674714331-yad-vashem-comes-to-minecraft-for-international-holocaust-day