Citizens infrastructures as a way to govern AI's power to shape our shared representations

How to make sure auditing institutions are aligned with values and societal expectations?

AI-generated images are one of the most glaring failures of auditing institutions in AI governance. Companies have been cultivating artificial image synthesis since the development of Generative Adversarial Networks (GAN) in 2014 (Elgammal et al. 2017), but since OpenAI's Generative Pre-trained Transformer (GPT) DALL-E, that generates images from text, was opened to the public in 2022 (OpenAI 2022), AI-generated images have flooded all public information and communication platforms (ICP) (Bond 2024). Presently, any internet image search displays some kind of AI-generated images, not all of which are immediately recognisable as such (Gangadharbatla 2022; Mazzone et al. 2019).

Yet, the public response to the large diffusion of AI-generated images is not at all a positive one[1]. The public laments the infestation by artificial images of spaces previously reserved to creative expression and inspiration such as Pinterest (Kimutai 2024), and the idea of AI generating visual art generates unnerving (Gray et al. 2012). Ratings of comfortableness towards and estimated capability of AI producing artistic images is very low compared to other areas of application of AI technology (Schepman et al. 2020), and the appreciation of AI-generated images is much lower than that of human crafted ones (Bellaiche et al. 2023; Hong et al. 2019; Chamberlain et al. 2017).

---

[1] A *caveat* to be introduced here, as most data available on the public's reaction to AI-generated images is culturally situated, and, while significant in its proportion for the purpose of calling out on current auditing processes in AI governance, doesn't necessarily represent all attitudes around the globe equally. See for example Wu et al. (2019) about the different perception of AI art among American test subjects as compared to Chinese.

The misalignment between the industry and the public is striking, but currently has no suitable reflection upon effective governance, as there is no adapted institution to assess, support and address this misalignment. With pervasive consequences. Indeed, ICPs where AI-generated images are taking up more and more space actually convey social and cultural cues that influence the construction of our shared beliefs and behaviours (Bail et al. 2018). Consequently, AI-generated contents themselves are already becoming a non-neglectable factor in our perception and appreciation of reality, influencing our opinions (Floridi 2024), our tastes (Wu et al. 2019) and our values (Brinkmann 2023), on a personal, interpersonal and societal level, with the potential to shape our perception of the self, our attitude towards others and our shared representations of the future, of democracy and of social identities (Ovadya et al. 2023). For such high stakes and wide-ranging consequences, we have no matching institutional levers: we lack infrastructures to connect between the big-tech industry, national and international regulatory instances, and the public. An issue for which I try to suggest a workable solution in this paper.

In fact, in the first part of my paper I show that the auditing institutions that have failed to govern the development of AI-generated images in alignment with societal values and expectations, have done so because of the lack of citizen infrastructures. These can take the shape of digital platforms, community forums, shared skills, information networks, or lived environments (AbdouMaliq 2004) and aim at supporting and enabling citizen-led and enduring public engagement (Gabrys 2019). Their specificity is that they often originate from everyday practices and in the very spaces where citizens interact (Livingstone et al. 2010).

I make the claim that such infrastructures can be particularly well-suited and effective for the purpose of enabling and supporting citizen engagement in auditing institutions involved in AI governance, effectively and durably aligning those institutions with social values and expectations. Indeed, a successful infrastructure connecting citizens and auditing institutions must allow governance issues to be formulated and discussed by citizens *contextually*, in a way that is coherent and organic with the live perception of these issues and with the experience-led evaluation of their consequences.

For this reason are citizens infrastructures so appealing and effective: they emerge from and in turn provide structure to the living environment or the everyday activities that are the very reason for citizens' perceptions, valuing and expectations regarding the issue at hand. For AI

regulation, this embedding of the infrastructure into the living experience would mean that the infrastructure is part of the citizens' consistent interaction with AI-generated content, which mostly happens through ICPs. This naturally points toward a kind of digital infrastructure, whose characteristics I outline in the last part of my paper, taking as case study the Interreg Europe project "CIVITEC", developed in partnership with the European Union with the objective of using digital tools to enable and support the alignment of citizen and institutions in governance debates. I analyse the propositions of CIVITEC to build a citizens infrastructure and discuss whether those are applicable in the case of AI governance.

Indeed, digital infrastructures have become more and more promising as technology advances (Barns 2016), but they have an especially interesting potential today to address the issue of aligning social values and expectations towards AI with auditing institutions. Indeed, the live context in which citizens encounter AI-generated content are ICPs, and ICPs can notably be the support for community forums, shared skills and information networks, which are the channels through which citizens infrastructures are commonly organised and maintained.

Hence, a digital citizens infrastructure appears to be a motivating option to govern AI's power to shape shared representations in alignment with social values and societal expectations. In fact, such a digital infrastructure would close the gap between consuming and governing AI-generated content, insofar as the space where AI-generated content has an impact on citizens' life and perception — ICPs— would also be the space where citizens can have an impact on the generation and diffusion of artificial content. This brings me to conclude my paper on the idea that it is possible to counter the trend that makes out people to be more passive as AI becomes more pervasive, and stress on the contrary how human-AI interaction can be a chance to enhance governance, interaction and decision-making.

List of references:

- AbdouMaliq, S. (2004) People as Infrastructure: Intersecting Fragments in Johannesburg. Public Culture (2004) 16 (3): 407–429. https://doi.org/10.1215/08992363-16-3-407

- Bail, C. et al. (2018) Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115.37 (2018), pp. 9216–9221.

- Barns, S., Cosgrave, E., Acuto, M., & Mcneill, D. (2016). Digital Infrastructures and Urban Governance. Urban Policy and Research, 35(1), 20–31. https://doi.org/10.1080/08111146.2016.1235032

- Bellaiche, L., Shahi, R., Turpin, M.H. et al. (2023) Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cogn. Research* 8, 42 (2023). https://doi.org/10.1186/s41235-023-00499-6

- Bond, S. (2004) AI-generated spam is starting to fill social media. Here's why. *NPR Untangling disinformation*. Accessed 27/02/2025. https://www.npr.org/2024/05/14/1251072726/ai-spam-images-facebook-linkedin-threads-meta

- Brinkmann, L., Baumann, F., Bonnefon, JF. et al. (2023) Machine culture. *Nat Hum Behav* 7, 1855–1868 (2023). https://doi.org/10.1038/s41562-023-01742-2

- Chamberlain, R., Mullin, C., Scheerlinck, B. and Wagemans, J. (2017) Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychol. Aesthet. Creat. Arts.*

- Elgammal, A., Liu, B., Elhoseiny, M., and Mazzone, M. (2017) CAN: Creative adversarial networks, generating "Art" by learning about styles and deviating from stylenorms. *arXiv* preprint. arXiv:1706.07068

- Luciano, F. (2024) Hypersuasion – On AI's Persuasive Power and How to Deal with It. Philos. Technol. 37, 64. https://doi.org/10.1007/s13347-024-00756-6

- Gabrys, J. (2019) Citizen Infrastructures and Public Policy: Activating the Democratic Potential of Infrastructures. In *Future Directions for Citizen Science and Public Policy*, edited by Katie Cohen and Robert Doubleday. Centre for Science and Policy (CSaP), University of Cambridge.

- Gangadharbatla, H. (2022). The Role of AI Attribution Knowledge in the Evaluation of Artwork. Empirical Studies of the Arts, 40(2), 125-142. https://doi.org/10.1177/0276237421994697

- Gray, K. and Wegner, D. M. (2012) Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition* 125 (1) pp. 125-130. https://doi.org/10.1016/j.cognition.2012.06.007

- Hong, J.-W. and Curran M. N. (2019) Artificial Intelligence, Artists, and Art: Attitudes Toward Artwork Produced by Humans vs. Artificial Intelligence. *ACM Transactions on*

Multimedia Computing, Communications, and Applications (TOMM) 15(2):58, 1 - 16. https://doi.org/10.1145/3326337

- Kimutai, L. (2024) Is AI killing Pinterest's magic? *Medium*. Accessed 27/02/2025. https://medium.com/@kimutail/is-ai-killing-pinterests-magic-2f84c395fd98

- Livingstone, S. and Lievrouw, L. A. (2010) How to infrastructure. In Star, S., & Bowker, G. (Eds.) *Handbook of New Media: Social Shaping and Social Consequences of ICT*, pp. 230-245. SAGE Publications Ltd, https://doi.org/10.4135/9781446211304

- Mazzone M. and Elgammal, A. (2019) Art, Creativity, and the Potential of Artificial Intelligence. *Arts* 8(1):26. https://doi.org/10.3390/arts8010026

- OpenAI (2022) DALL·E 2. Accessed 27/02/2025. https://openai.com/index/dall-e-2/

- Ovadya, A. and Thorburn, L. (2023) Bridging Systems. Open Problems for Countering Destructive Divisiveness across Ranking, Recommenders, and Governance. 23-11 *Knight First Amend. Inst.* (Oct. 26, 2023), https://knightcolumbia.org/content/bridging-systems

- Schepman, A. and Rodway, P. (2020) Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports* 2020(1). https://doi.org/10.1016/j.chbr.2020.100014

- Wu, Y., Yi Mou, Zhipeng Li, Kun Xu (2019) Investigating American and Chinese Subjects' explicit and implicit perceptions of AI-Generated artistic work. *Computers in Human Behavior* 2019(104). https://doi.org/10.1016/j.chb.2019.106186.