

zde

Zentrum für
Demokratie
Aarau

Center for Research on Direct Democracy | UZH
Allgemeine Demokratieforschung | UZH
Politische Bildung und Geschichtsdidaktik | PH FHNW

Digital citizenship education in the age of AI

Manuel S. Hubacher

Aarauer Demokratietage 2026, Democracy without trust?

13 March 2026

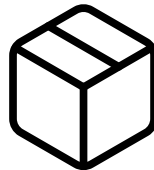
Three Challenges



Disinformation

Undermines trust in democratic institutions

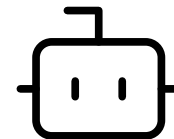
(e.g., Chesney and Citron 2019; Hubacher 2021; Mansell et al. 2025)



AI as Black Box

Opaque decisions, algorithmic discrimination

(e.g., Al-Amoudi et al. 2019; van Dis et al. 2023; von Eschenbach 2021)



Solutionism & Determinism

Fixing AI problems with more AI instead of addressing the challenges

(e.g., Holmes and Tuomi 2022; Jandrić und Knox 2022; Mühlhoff 2025)

Critical Thinking vs. Trust?



**Critical thinking/
scepticism**

modus operandi for
democratic citizens?

VS.



Trust

condition for democracy?

Thesis:

Digital citizenship education for
the age of AI should foster
epistemic vigilance and
transformative capabilities



Trust

Trusting φ means that we accept φ 's utterances or outputs O_φ (e.g., φ 's advice) without questioning O_φ .

Trust can be understood as a **non-critical** and **accepting attitude**.

(Hauswald, 2023; Nguyen 2023)



Critical thinking

A mix of higher-order reasoning, self-reflection, and attention to power, justice, and participation.

Scepticism is a core disposition within critical thinking.

(Andreucci-Annunziata et al. 2023; Bensley 2023; Hitchcock 2024)

Healthy amount of scepticism



Too much

Weaponised 'scepticism'/ unjustified mistrust: destroying the epistemic minimal consensus that democratic deliberation requires.

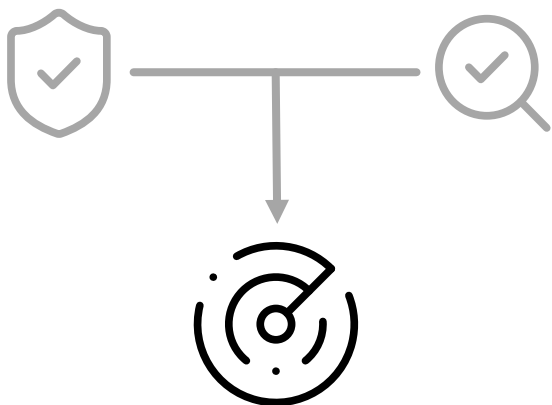
(e.g., McKay and Tenove 2021)



Too little

Uncritical deference to authority undermines democratic agency. Danger of easier manipulation.

(e.g., Bertou 2019; McKay and Tenove 2021)



Epistemic vigilance

(Hauswald, 2023; Nguyen 2023)

1

‘Direct’ vigilance

Evaluating individual media contents, arguments, and claims – reading competence, analytical skills, domain knowledge.

2

‘Indirect’ vigilance

Assessing the trustworthiness of institutions and sources as such – political knowledge, understanding of media ecosystems, epistemic capacities.

Critical media literacy +

Critical media literacy

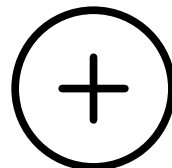
Function: Navigate the status quo

Artefact

'Direct' vigilance: Critical engagement with media content, identify bias and manipulation

System

'Indirect' vigilance: Assess institutional trustworthiness; contextualise sources



Critical-transformative engagement

Function: Emancipate from the status quo

What kind of artefacts (e.g. AI systems) do we need?

How should a system (political system, media system, etc.) be designed in order to earn trust?

Stories of the Future

How could this look like?

SNSF Agora Project | Geneva Graduate Institute | 2022–2024



30+ narrative foresight workshops across Switzerland with youth as co-researcher
(age of participants: 15–25)



Critical examination of AI's role in democratic processes



Participatory construction of desirable futures through storytelling (and art)

(Dubbery et al. 2024; Hubacher et al. 2024)

Conclusion

- 1 Citizens must trust trustworthy institutions: but that trust must remain responsive, reactivating critical scrutiny when warranted.
- 2 Digital citizenship education should foster epistemic vigilance and capabilities for critical-transformative engagement.
- 3 This requires analytical and critical competencies, future literacy, agency, and systemic and technical knowledge – both on the part of citizens and teachers.

References

- Andreucci-Annunziata, Paola, Andrea Riedemann, Susana Cortés, Augusto Mellado, María Teresa del Río, and Alejandro Vega-Muñoz. 2023. 'Conceptualizations and Instructional Strategies on Critical Thinking in Higher Education: A Systematic Review of Systematic Reviews'. *Frontiers in Education* 8 (March). <https://doi.org/10.3389/educ.2023.1141686>.
- Bensley, D. Alan. 2023. 'Critical Thinking, Intelligence, and Unsubstantiated Beliefs: An Integrative Review'. *Journal of Intelligence* 11 (11). <https://doi.org/10.3390/jintelligence11110207>.
- Bertsou, Eri. 2019. 'Rethinking Political Distrust'. *European Political Science Review* 11 (2): 213–30. <https://doi.org/10.1017/S1755773919000080>.
- Chesney, Bobby, and Danielle Citron. 2019. 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security'. *Calif. L. Rev.* 107: 1753. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/calr107§ion=51.
- Dis, Eva A. M. van, Johan Bollen, Willem Zuidema, Robert van Rooij, and Claudi L. Bockting. 2023. 'ChatGPT: Five Priorities for Research'. *Nature* 614 (7947): 224–26. <https://doi.org/10.1038/d41586-023-00288-7>.
- Dubbery, Jérôme, Christine Lutringer, Laura Bullon-Cassis, and Manuel Hubacher. 2024. 'Une gouvernance mondiale de l'IA est-elle possible et si oui, peut-on y intégrer la jeunesse?' The Conversation, May 27. <http://theconversation.com/une-gouvernance-mondiale-de-lia-est-elle-possible-et-si-oui-peut-on-y-integrer-la-jeunesse-230803>.
- Eschenbach, Warren J. von. 2021. 'Transparency and the Black Box Problem: Why We Do Not Trust AI'. *Philosophy & Technology* 34 (4): 1607–22. <https://doi.org/10.1007/s13347-021-00477-0>.
- Hauswald, Rico. 2024. 'Caveat usor: Vertrauen und epistemische Wachsamkeit gegenüber künstlicher Intelligenz'. *Zeitschrift für Praktische Philosophie* 11 (1): 367–94. <https://doi.org/10.22613/zfpp/11.1.15>.
- Holmes, Wayne, and Ilkka Tuomi. 2022. 'State of the Art and Practice in AI in Education'. *European Journal of Education* 57 (4): 542–70. <https://doi.org/10.1111/ejed.12533>.
- Hubacher, Manuel S. 2021. 'Sie sind Fake News! Ein analytischer Zugang für die Politische Bildung'. In *Politische Bildung für die digitale Öffentlichkeit: Umgang mit politischer Information und Kommunikation in digitalen Räumen*, edited by Manuel S. Hubacher and Monika Waldis. Springer VS. https://doi.org/10.1007/978-3-658-33255-6_8.
- Hubacher, Manuel S., Laura Bullon-Cassis, Jérôme Duberry, and Christine Lutringer. 2024. 'Stories of the Future: Mit Geschichten mögliche Zukünfte erforschen – ein Praxisbericht'. *POLIS*. <https://doi.org/10.26041/fhnw-10511>.
- Jandrić, Petar, and Jeremy Knox. 2022. 'The Postdigital Turn: Philosophy, Education, Research'. *Policy Futures in Education* 20 (7): 780–95. <https://doi.org/10.1177/14782103211062713>.
- Mansell, Robin, Flavia Durach, Matthias Kettemann, et al. 2025. *Information Ecosystem and Troubled Democracy: A Global Synthesis of the State of Knowledge on News Media, AI and Data Governance*. Observatory on Information and Democracy. <https://observatory.informationanddemocracy.org/report/information-ecosystem-and-troubled-democracy/>.
- McKay, Spencer, and Chris Tenove. 2021. 'Disinformation as a Threat to Deliberative Democracy'. *Political Research Quarterly* 74 (3): 703–17. <https://doi.org/10.1177/1065912920938143>.
- Mühlhoff, Rainer. 2025. *Künstliche Intelligenz und der neue Faschismus*. Reclams Universal-Bibliothek 962455. Reclam.
- Nguyen, C. Thi. 2022. 'Trust as an Unquestioning Attitude'. *Oxford Studies in Epistemology* 7: 214–44.
- *ZDF Magazin Royale*. 2026. Season 12, Episode 2, 'Willkommen im Epstein-Rabbit-Hole'. ZDF, Juni 3. <https://www.zdf.de/video/shows/zdf-magazin-royale-102/zdf-magazin-royale-vom-6-maerz-2026-100>.