1. Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2022). **Deepfake detection by human crowds, machines, and machine-informed crowds**. *Proceedings of the National Academy of Sciences*, *119*(1), e2110013119. – Exploration of the cognitive processes involved in Deep Fake detection by Humans. In this Study, participants were given pairs of videos (one real and one deepfake) and asked to find the deepfake video. And also, participants were asked to rate their confidence about their findings whether the chosen one is real or deep fake. Participants were given chance to correct their answers after showing them the AI model’s prediction where it refers to the Cognitive reflection rate (not mentioned in paper).
2. Hashmi, A., Shahzad, S. A., Lin, C. W., Tsao, Y., & Wang, H. M. (2024). **Unmasking Illusions: Understanding Human Perception of Audiovisual Deepfakes**. *arXiv preprint arXiv:2405.04097*. – This study is comparison on human perception and AI detection models on audiovisual deepfakes. Each participant watched 40 videos twice in different random orders. While the experiment, first they show all real and deepfake videos without label to find them, and after they find, then the correct labels were displayed which one is real or fake. Thus, they predicted the participants performance. Humans are not that good in their detection performance in this study where the study found there is an influence of participants’ demographics like age, native language, and IT skill level. And also evaluated the correlation between their self-reported confidence and actual accuracy.
3. Frank, J., Herbert, F., Ricker, J., Schönherr, L., Eisenhofer, T., Fischer, A., ... & Holz, T. (2024, May). **A representative study on human detection of artificially generated media across countries**. In *2024 IEEE Symposium on Security and Privacy (SP)* (pp. 55-73). IEEE. -
4. Lewis, A., Vu, P., Duch, R. M., & Chowdhury, A. (2023). **Deepfake detection with and without content warnings.** *Royal Society Open Science*, *10*(11), 231214. – This study’s aim is to find people’s ability to detect deepfake videos with and without content warnings. With having three set of group participants where first one is a control group watching only real vides, second one is viewing one deepfake video among the real videos without warning and the latter one is viewing the same videos but with warned that at least one was fake. Important findings show that without warnings, participants who watch a deepfake were not able to find anything unusual compared to those who only view real ones. Even with warning, only twenty percent of the participants correctly identify the deepfake as the only fake video, whereas nearly half mistakenly selected one or more real videos as deepfakes. The author concluded that effective identification, management, and mitigation of deepfake content may require trust in external authentication sources, as content warnings alone may not make individuals to accurately detect deepfakes.
5. Somoray, K., & Miller, D. J. (2023). **Providing detection strategies to improve human detection of deepfakes: An experimental study**. Computers in Human Behavior, 149, 107917. – The study involved participants giving 20 videos as both real and deepfake. Participants were randomly assigned to get deepfake “detection strategies” (clues for AI generated) or a control group. The participants’ detection accuracy and confidence were measured in the research study. There was a small positive correlation between these both measurements. The detection strategies didn’t help to impact accuracy or confidence, no bias was found towards categorizing videos as real.
6. Halpin, S. N. (2024). Inter-Coder Agreement in Qualitative Coding: Considerations for its Use. American Journal of Qualitative Research, 8(3), 23-43. - to calculate cohen kappa inter-coder agreement score.
7. Different researchers, different results? analyzing the influence of researcher experience and data type during qualitative analysis of an interview and Survey Study on Security Advice. (n.d.). <https://dl.acm.org/doi/fullHtml/10.1145/3544548.3580766>

* read this paper for referring to preparing a codebook

1. M. Appeland F. Prietzel, “**The detection of political deepfakes**,” JournalofComputer-MediatedCommunication,2022. - Studies on the detection of political deepfakes, which may involve aspects of human concentration and cognitive reflection.
2. G. Pennycook and D. G. Rand, “**Who Falls for Fake News? The Roles of Bullshit Receptivity, Overclaiming, Familiarity, and Analytic Thinking,**” Journal of Personality,2020. - Research on who falls for fake news, suggests that susceptibility to fake news is better r told by the lack of reasoning rather than motivated reasoning, which could be related to concentration and attention of humans.
3. S.-H.Jeong, H.Cho, and Y.Hwang, “**Media Literacy Interventions: A Meta-Analytic Review**, ” Journal of Communication, 2012. - A meta-analytic study indicating that people with a better understanding of media and media production systems are more skeptical and realistic about media messages. Experience makes detection more accurate.
4. Ortloff, A. M., Fassl, M., Ponticello, A., Martius, F., Mertens, A., Krombholz, K., & Smith, M. (2023, April). Different researchers, different results? analyzing the influence of researcher experience and data type during qualitative analysis of an interview and survey study on security advice. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-21). – Qualitative data analysis, coder duties, codebook creation, policies
5. S. J. Nightingale and H. Farid, “**AI-synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy**,” Proceedings of the National Academy of Sciences, 2022. - A study on human visual perception of artificial faces, showing that AI generated faces are same from real faces and are considered as more trustworthy.
6. S. Ahmed, “**Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes**,” Personality and Individual Differences, 2021. - A study on cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes.
7. Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). **The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings.** *Management science*, *66*(11), 4944-4957.
8. Chan, M. P. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). **Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation**. *Psychological science*, *28*(11), 1531-1546. - The study concludes that effective debunking requires more than merely labeling information as false.

Optional:

1. Bialek, M., & Pennycook, G. (2018). **The cognitive reflection test is robust to multiple exposures**. *Behavior research methods*, *50*, 1953-1959.
2. Abraham, J., Putra, H. A., Prayoga, T., Warnars, H. L. H. S., Manurung, R. H., & Nainggolan, T. (2022). **Prediction of self-efficacy in recognizing deepfakes based on personality traits**. *F1000Research*, *11*.

[**Extra Info:** Deepfake studies, such as those involved in the FaceForensics++, Celeb-DF and FakeAVCeleb.]

Google Drive:  
Celeb-DF (v2) dataset:   
[https://drive.google.com/open?id=1iLx76wsbi9itnkxSqz9BVBl4ZvnbIazj](https://www.google.com/url?q=https://drive.google.com/open?id%3D1iLx76wsbi9itnkxSqz9BVBl4ZvnbIazj&sa=D&source=editors&ust=1730910653529219&usg=AOvVaw3lrlUzqMw8vQyUtyK9Y5NP).   
  
Celeb-DF (v1) dataset:   
[https://drive.google.com/open?id=10NGF38RgF8FZneKOuCOdRIsPzpC7\_WDd](https://www.google.com/url?q=https://drive.google.com/open?id%3D10NGF38RgF8FZneKOuCOdRIsPzpC7_WDd&sa=D&source=editors&ust=1730910653529375&usg=AOvVaw1FPD2RBgPfjoyC8vaQOLYI).   
  
Baidu Net Disk:  
Celeb-DF (v2) dataset:   
[https://pan.baidu.com/s/1EcYX0s4U3kbI1V2vdrP46A](https://www.google.com/url?q=https://pan.baidu.com/s/1EcYX0s4U3kbI1V2vdrP46A&sa=D&source=editors&ust=1730910653529444&usg=AOvVaw0q3lPypJUqmWlz0ilbgFKd)   
code：yxa1   
  
Celeb-DF (v1) dataset:   
[https://pan.baidu.com/s/16QulfMFG4TQB9iMZIZnsjQ](https://www.google.com/url?q=https://pan.baidu.com/s/16QulfMFG4TQB9iMZIZnsjQ&sa=D&source=editors&ust=1730910653529515&usg=AOvVaw1y5Ns8E3YrNIgz0YRq07-I)   
code：ku0s

Audio-Video Deepfake dataset (FakeAVCeleb)

<https://sites.google.com/view/fakeavcelebdash-lab/>

<https://github.com/DASH-Lab/FakeAVCeleb/tree/main/dataset> - Fill in the Google form mentioned in this link to get the deepfake resources same as I got for the he above Celeb-DF datasets.