**Data Analysis Plan**

**Data-preparation:**

1. The data of participants who do not fully complete all questions and tasks will be excluded from analyses during the first round of analyses.

2. The data will be excluded of participants who had pIAT error rates for any of the pIATs above 30% across the entire task, or above 40% for any one of the four critical blocks or for participants who complete more than 10% of pIAT trials faster than 400 ms.

3. The D2 algorithm will be used to create pIAT scores. Scores will be calculated so that positive values reflected a relative response bias for Chris over Bob whereas negative values indicated the reverse pattern of responding (a relative response bias favoring Bob over Chris).

**Data-analyses:**

**Video content**. An average self-reported rating score for Chris will be calculated by averaging responses from the three Likert rating scales. This mean score will be submitted to an independent samples t-test with *video content* (Positive vs. Negative) as a between subjects factor. pIAT scores will be submitted to a similar set of analyses. In addition we will also carry out a single sample t-test to examine if self-reported and pIAT scores differ from zero, one for those in the positive content condition and another for those in the negative content condition. In all cases, effect sizes (Cohen’s d) will be reported. We will also compute Bayesian factors in accordance with procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009) to estimate the amount of evidence for the hypothesis that stimulus evaluations differ as a function of video content (alternative hypothesis) or that there is no difference (null hypothesis).

**Video type**. An independent samples t-test will be carried out in order to examine if the genuine and Deepfaked videos differ in the evaluations that they produce. Data will first be recoded so that the valence of the video content is controlled for (i.e., scores from those in the negative content groups will be re-coded by multiplying their values by -1). Effect sizes (Cohen’s d) will be reported. We will also compute Bayesian factors in accordance with procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009) to estimate the amount of evidence that stimulus evaluations differ as a function of video type (alternative hypothesis) or that there is no difference (null hypothesis).

**Demographic and Individual Difference Measures**

**Demographic variables**. Age, gender, ethnicity, and location of residence variables will not be recoded. Those who provide household income values will be assigned a number ranging from 1 (*lower income*: Less than $25,000) to 8 (*higher income*: $200,000 or more). Participants who provided a response on the employment status question will be grouped in the following ways: those who respond ‘Employed for wages (part-time)’, ‘Employed for wages (full time)’, or ‘Self-employed’ will be assigned to one group (‘Currently Employed’); those who respond ‘Out of work and looking for work’, ‘Out of work and not looking for work’, ‘Retired’, or ‘Unable to work’, will be assigned to a second group (‘Not Currently Employed’) while those who responded ‘A homemaker’, ‘A student’, or ‘military’ will be assigned to a third group (‘Vocational Occupation’). Participants will be assigned a ‘General Education’ score ranging from 0 (the lowest educated group: Less than a high school degree) to 7 (the highest educated group: Doctoral degree or Professional degree (JD, MD)).

**Individual Differences**

**Political Ideology**. Participants will be assigned two separate political ideology scores each ranging from 1 (*Strongly Liberal*) to 5 (*Strongly Conservative*) – one for their ideology concerning economic issues and another for social issues. If these two scores are found to correlate highly with one another then we will average them to create a ‘General Political Ideology’ score, with higher values indicative of more conservative beliefs and lower values of more liberal beliefs.

**Political Identity**. Participants will be asked two questions concerning their political identity. Their responses will be scored from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*). If these two scores are found to correlate highly then we will average them to create a ‘General Political Identity’ score, with larger values indicating greater endorsement of the idea that one’s political attitudes and beliefs are important to one’s self-identity.

**Religious Beliefs**. Participants will be asked to indicate their religious affiliation. We will then classify them based on their response into one of three groups: Religious, Agnostic, Atheist. Participants responses on the Religious Belief Scale can range from 1 (*I strongly disagree*) to 5 (*I strongly agree*). Responses to each item will be summed to create a total score representing ‘Religiosity’, with higher scores indicating higher religiosity.

**Cognitive Ability**. A sum score will be created based on the number of items correctly answered in the Revised Cognitive Reflection test (number of items correct out of 7).

**Cognitive Thinking Disposition**. For scoring the ‘Cognitive Thinking Disposition’, we will first reverse the appropriate items from the Rational-Experiential Inventory, and then calculate the mean of the 10 items measuring Rationality and the mean of the 10 items measuring Experientiality.

**Conspiratorial Thinking**. An overall score will be computed based on the mean of all Belief in Conspiracy Theory Inventory items, with higher scores reflecting stronger belief in conspiracy theories and lower scores the opposite.

**Overclaiming**. Paulhus et al. (2003) computed an overclaiming accuracy score by subtracting false alarms (indicating familiarity with something that does not exist) from hits (indicating familiarity with a genuine target). For ease of exposition, we simply reversed this equation so that a higher score indicates more overclaiming (i.e., a higher incidence of reporting impossible knowledge relative to actual knowledge). Results for the overclaiming measure are similar if false alarms are used as the primary measure instead of computing the overall accuracy score.

**Analyses**. We will correlate self-reported evaluations, pIAT scores, and deepfake detection scores with the aforementioned individual difference and demographic measures.