Deep-learning-based face-swap video and audio manipulations (Deepfakes) are getting easier to make, reducing trust in media. This project, DeFake, aims to develop a tool to help journalists in the efficient detection of these manipulations before they reach the general population. We performed user studies with 11 journalists and media verification workers to develop a prototype and identify the requirements of a deepfake detection tool.

**Requirements for Journalists**

Providing explainability.

- Variety of distinct detection methods to allow journalists make informed judgements.
- Time-distributed fakeness scores.
- Visualization of manipulated areas/faces.

**Technical Requirements**

- Accuracy over speed
- Less false negatives
- Open Source
- Logins and rate limiting

Accuracy is the most important thing in journalism. Anyone who tells you otherwise is not a good journalist.

---

**Current Process**

Media verification triggered by: bipartisan sources and polarizing content.

- No deepfake detection tools exist.
- Low awareness.
- Underestimating quality of manipulations.

**Expected Workflow**

The tool to be used for supplementary information.

- Receive Video
- Trusted?
- Compile Story
- DeFake Tool
- Talking Head?
- Context Analysis

**Moving Forward**

Beta testing a deployed tool on journalists for iterative improvements to the interface & workflow.