**Motivation**
- Filter rules power privacy-enhancing technologies (PET) to block ads and tracking, but they are still maintained by human experts.
- Researchers utilize machine-learning (ML) approaches to replace or assist filter rule creation. Yet, they rely on existing rules to label their ground truth for ML models, creating a circular dependency.

**AutoFR Contributions**
- **Framework:** Using reinforcement learning (RL) to automate URL-based rule generation that block ads without using existing rules.
- **User preference (w):** A knob to express user preference of avoiding visual breakage (missing images/text), automatically.
- **Practical and Scalable:** AutoFR is a tool that takes 1.6 minutes per-site and can scale to millions of sites and over time.

**Reinforcement Learning Framework**
- **Environment**
  - User of AutoFR
  - Site [0]
  - Browser
  - URL
  - Optimize: $	ext{B} = \frac{\hat{C}_T + \hat{C}_I}{2}$

- **Agent**
  - Config
  - Filter Rules
  - Action (Filter Rule)
  - Policy
  - Output

- **Action Space**
  - Updates
  - Action Space
  - Filter Rules

- **Breakage**
  - Effectiveness of a Filter Rule: Is measured by the proportion of ads ($C_A$) that were removed (good) vs. how much legitimate content, such as images ($C_L$) and text ($C_T$), went missing (bad), when applying the rule on a site. Threshold $w$ is a user given preference to denote how much the user cares about avoiding breakage ($B$).

  $B = \frac{\hat{C}_I + \hat{C}_T}{2}$

- **Reward**
  - $R_f(w, C_A, B) = \left\{ \begin{array}{ll}
  -1 & \text{if } \hat{C}_A = 0 \\
  0 & \text{if } \hat{C}_A > 0, 1 - B < w \\
  \hat{C}_A & \text{if } \hat{C}_A > 0, 1 - B \geq w
  \end{array} \right.$

**Evaluation: AutoFR vs. EasyList**
- **Performance**
  - Blocking Ads ($\hat{C}_I$)
  - Avoiding Breakage ($1 - \hat{B}$)

  - (b) AutoFR (In the Wild)
  - (c) EasyList (In the Wild)

- **86% of sites (AutoFR) vs. 87% (EasyList):** We apply AutoFR to the Top-5K websites in the wild and find that it achieves comparable results for blocking ads while avoiding visual breakage within the given threshold ($w=0.9$) to EasyList, the state-of-the-art filter list used by PET.

- **361 Rules Generated:** 70% of the Top-20 rules generated by AutoFR appear in EasyList (EL). 40% of these rules match with the Top-20 rules from EasyList.

**Broader Impact**
- **Ad-Filtering Dev Summit:** Hieu Le presented AutoFR to the PET community on Oct. 2022 and is working closely with the community to get AutoFR adopted by industry players.

**Future Directions**
- **Other Platforms:** Extend AutoFR to platforms such as mobile, smart TVs and Oculus VR.
- **Tracking and Functionality:** Create rules to block tracking and avoid functionality breakage (forms, scrolling).