



# Frühwarnsystem für Falschmeldungen

## Ziel

Das Ziel von Manuel Gomez Rodriguez vom Max-Planck-Institut für Software ist es, die häufig im Netz verbreiteten Falschmeldungen durch einen hybriden Ansatz aus künstlicher Intelligenz und menschlichen Bewertungen effektiv aufzudecken, sowie Debatten im Internet zu versachlichen.

## Fake-News

Gefahr: Sie polarisieren und manipulieren

## Algorithmen

### Curb:

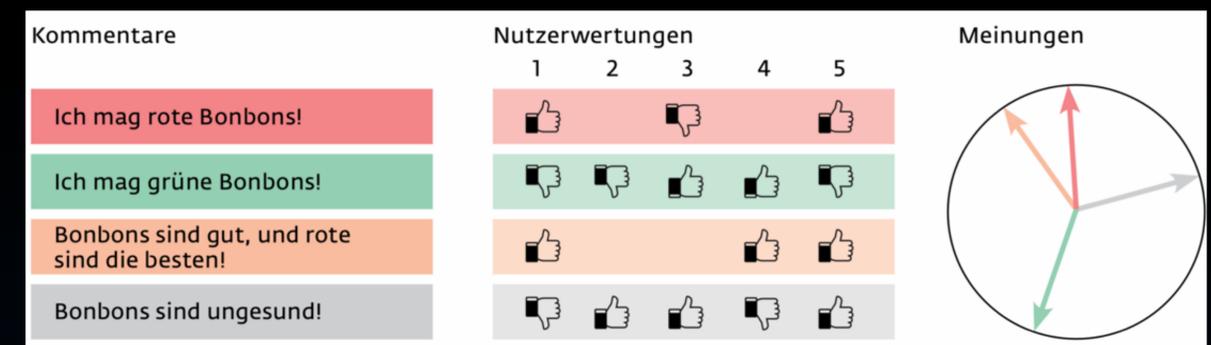
Curb ist ein von Manuel Gomez Rodriguez entworfener Algorithmus, der priorisiert, welche Inhalte die nur begrenzt verfügbaren menschlichen Faktenchecker am dringendsten überprüfen und gegebenenfalls als falsch melden müssen. Dabei nimmt der Algorithmus in den Blick, wie schnell sich eine Meldung verbreitet und wie oft sie als falsch deklariert wird. Ziel dabei ist es, eine optimale Balance zwischen dem Bemühen, möglichst wenig Menschen mit der undeckelten Falschmeldung zu konfrontieren und der Effizienz beim Einsatz der menschlichen Faktenchecker, zu erreichen.

### Zusätzliche Anwendungsbereiche

- Optimieren von Sprachlern-Software
- Virales Marketing
- Curb wurde ursprünglich dafür entwickelt

### Detective:

Der Algorithmus Detective soll ebenfalls die Verbreitung von Falschmeldungen verhindern. Da bei dem Algorithmus Curb das Problem besteht, dass Nutzer das System gezielt sabotieren können, haben Rodriguez und sein Team „Detective“ entwickelt. Der Detective-Algorithmus, der nicht so wie Curb alle Nutzer für gleich seriös hält, versucht herauszufinden, wer Fake-Meldungen besonders zuverlässig als Fake deklariert und wer solide Nachrichten vorsätzlich als Fake brandmarkt. Praktisch setzt da aber der Datenschutz Grenzen, weshalb die Resultate von Detective entsprechend anonymisiert werden müssen.



## Test mit echten Daten

- Bestand aus Datensätzen von Twitter und Weibo
- Trat gegen das Pseudoverfahren Oracle, welches im Testszenario Zugang zu den Informationen hatte, ob eine Meldung Fake war oder nicht, und gegen drei andere Vergleichsmethoden basierend auf einfachen Faustregeln an
- Curb verhinderte fast genau so gut wie Oracle die Verbreitung von Falschmeldungen, die drei Faustregeln vermochten dies nicht
- Detective bestand, ähnlich wie Curb, den Test mit Bravour