



Bundesministerium  
für Bildung  
und Forschung



Freie Universität Berlin

# NOHATE

„...das ist halt ein schmaler Grad manchmal...“  
- Moderationspraxis in deutschen Redaktionen -

No Hate Speech-Kongress, Friedrich-Ebert-Stiftung, 8. April 2019, Berlin

Sünje Paasch-Colberg, Freie Universität Berlin

# NOHATE: BMBF-Verbundprojekt



- BMBF-Förderlinie „Zusammenhalt stärken in Zeiten von Krisen und Umbrüchen“
- Verbundpartner:
  - Freie Universität Berlin: Institut für Publizistik- und Kommunikationswissenschaft
  - Beuth Hochschule für Technik Berlin: FB Informatik und Medien
  - VICO Research & Consulting GmbH
- Unterstützung & Kooperation:
  - Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration
  - Amadeu Antonio Stiftung
  - Neue deutsche Medienmacher e. V.
  - Medienpartner (WELT, Heise, gutefrage.net)

# NOHATE: Projektziele

- Identifikation, Kategorisierung und Analyse (der Ursachen & Dynamiken) von Hasskommentaren im Internet
- Automatisierte Identifikation von Hate Speech im Themenfeld Flucht und Migration
- Softwaregestützte Anwendung zur Unterstützung der Content-Moderation



# Qualitative Interviewstudie



**NO**HATE

# Studiensteckbrief

## Methode

- 20 Leitfadeninterviews (~50 min)
- Community Manager\*innen & Content Moderator\*innen von Nachrichtenwebseiten, Radio, TV, Blogs und Foren
- Insgesamt knapp 30 Fragen zu
  - Arbeitsbedingungen & Selbstverständnis
  - Kommentare & Diskussionen im Allgemeinen
  - Hasskommentare & Moderationsstrategien

## Kontext: Arbeitsbedingungen

- Unterschiedliche Einbindung des CM in die Redaktionsstrukturen
- Tätigkeitsbereiche sehr divers
- Häufig fehlende Ressourcen
- Z. T. fehlende Unterstützung durch Redaktion
- Netiquette als zentrale Orientierung, sonst hohe Selbstständigkeit in der Praxis

# Ergebnisse: Hasskommentare

## Verständnis „Hasskommentar“

- Selten: Eindeutige Fälle
  - Gewalt-/Morddrohungen, Beleidigungen
  - Rassismus, Sexismus, Antisemitismus
- Häufig: Implizites, Codiertes, „Grauzonen“
  - Negative Verallgemeinerung durch Kontext
  - Herabwürdigung als Intention erkennbar
- Unterschiedliche „Schmerzgrenzen“

➤ **Inhalt, Kontext, Intention, Wirkung**

## Ausdrucksformen

- Schimpfwörter, Entmenschlichungen, negative Pauschalisierungen
- Bedrohungsszenarien
- „Tarnstrategien“ und codierte Sprache:
  - Ironische Wortschöpfungen („Goldstücke“)
  - Verballhornung von Namen („Angela Ferkel“)
  - Rhetorische Fragen
  - Leetspeak
  - Unpassende Emoticons

# Ergebnisse: Hasskommentare

## Verständnis „Hasskommentar“

- Selten: Eindeutige Fälle
  - Gewalt-/Morddrohungen, Beleidigungen
  - Rassismus, Sexismus, Antisemitismus
- Häufig: Implizites, Codiertes, „Grauzonen“
  - Negative Verallgemeinerung durch Kontext
  - Herabwürdigung als Intention erkennbar
- Unterschiedliche „Schmerzgrenzen“

➤ **Inhalt, Kontext, Intention, Wirkung**

## Ausdrucksformen

- Negative Verallgemeinerungen als Intention:
  - Posten eines Links
  - Posten eines mutmaßlichen Beweises

# Ergebnisse: Hasskommentare

## Ursachen?

- Thema als Hauptursache: Flucht/Migration, Islam/Religion, Kriminalität, Gender/Feminismus, Fußball, Impfen/Homöopathie, Rente, Kinderbetreuung
- Journalistische Darstellung: Emotionalisierung
- Bestimmte Personen: Angela Merkel, junge Politiker\*innen, B90/Die Grünen
- Bestimmte Charakteristika der Autor\*innen: Junge Frauen, mutmaßlich nicht-deutsche Namen

## Dynamiken?

- Typische Muster der Eskalation:
  - Hasskommentare kommen schnell
  - Aggressiver Kommentar als Auslöser
  - Eskalationsdynamik: „hochschaukeln“, „anschwellen“
- Aber: Grenzwertiger Kommentar auch Auslöser von Gegenrede



# Ergebnisse: Moderationsstrategien

## 1 - Vermeidungsstrategie

- Beschränkter Zugang zur Diskussion
  - z. B. nur Abonnet\*innen
  - z. B. keine Kommentare bei YouTube
- Kommentarmöglichkeit nur für bestimmte Themen oder Zeiträume

## 2 - „Die Masse beherrschen“

- Offene Diskussion & „unsichtbare“ Moderation
  - Tools: Wortfilter, Flagging, User-Monitoring
  - Kontrolle von Themen und Timing
  - Fokus auf sensible Themen: Präsenz der Moderation, frühe sichtbare Intervention, strengeres Löschen/Blockieren
  - Fokus auf sensible Tageszeiten

# Ergebnisse: Moderationsstrategien

## 3 – Positives fördern

- Fokus auf wertvolle Kommentare
  - Wertschätzung konstruktiver Kommentare: Hervorhebung als „editor’s pick“
  - Einbindung von Kommentaren in Beiträge
- Sonderformate der direkten Begegnung
  - z. B. Diskussion per Video-Stream



# Thesen: Gegenstrategien

- Definition(en):
  - Heterogenes Phänomen = verschiedene Gegenstrategien
- Ziele:
  - Für „Mitlesende“ Gegenrede machen
- Inhalte:
  - Über typische Hass-Rhetorik aufklären
  - Gerüchte und Falschmeldungen entkräften
- Kontextfaktoren/Auslöser:
  - Hintergrund/Netzwerk von Hatern offenlegen
  - Keine Deutungshoheit zulassen: Präsenz zeigen, alternative Angebote & Argumente schaffen
  - Emotionalisierende Überschriften vermeiden
  - Ressourcen konzentrieren
- Dynamiken:
  - Grenzwertiges kann auch Gegenrede auslösen

# Thesen: Gegenstrategien

- Gegenmaßnahmen:
  - Konstruktive Beiträge wertschätzen
  - Vorsicht mit Ironie & Humor
  - Technische Möglichkeiten stärker nutzen
- Infrastruktur:
  - Bedarf an Austausch, Training und Guidelines
  - Strukturen, die Erfahrungen mit Hate Speech ernst nehmen

nohate.online

s.colberg@fu-berlin.de  
@paaschcolberg

christian.strippel@fu-berlin.de  
@strippel

martin.emmer@fu-berlin.de  
@mjemmer

joachim.trebbe@fu-berlin.de  
@vielseher

Vielen Dank!

**NO**HATE