The Radicalization Pipeline:

Social Media, Extremism, and AI Countermeasures

PRESENTATION BY:
NICOLAS STEPHENS

Table of Contents

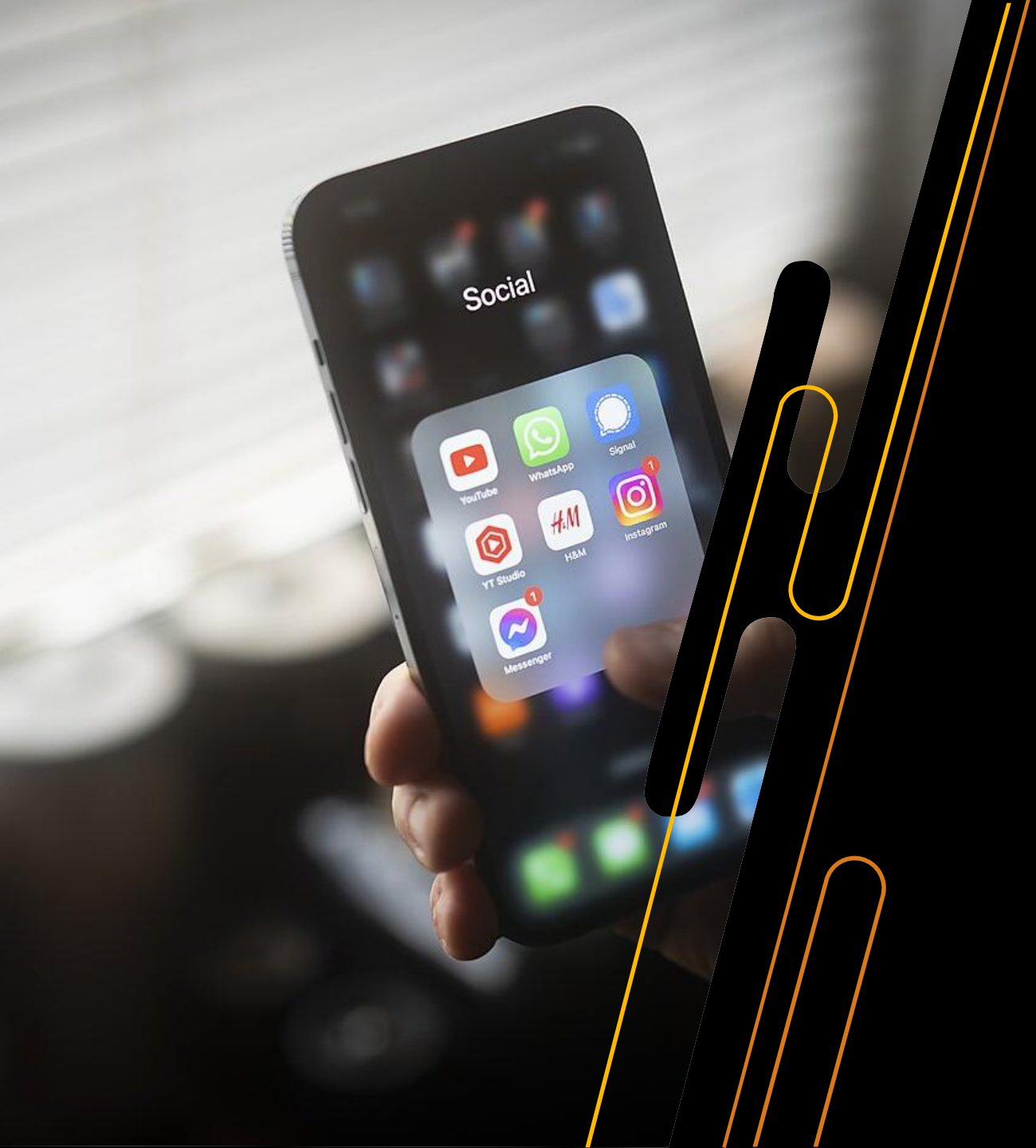
01. How Social Media Spreads Extremist Content

02. The Online Radicalization Process

03. Social Sciences Principles Behind Online Radicalization

04. LLM-Based Detection Techniques

05. Ethical Considerations



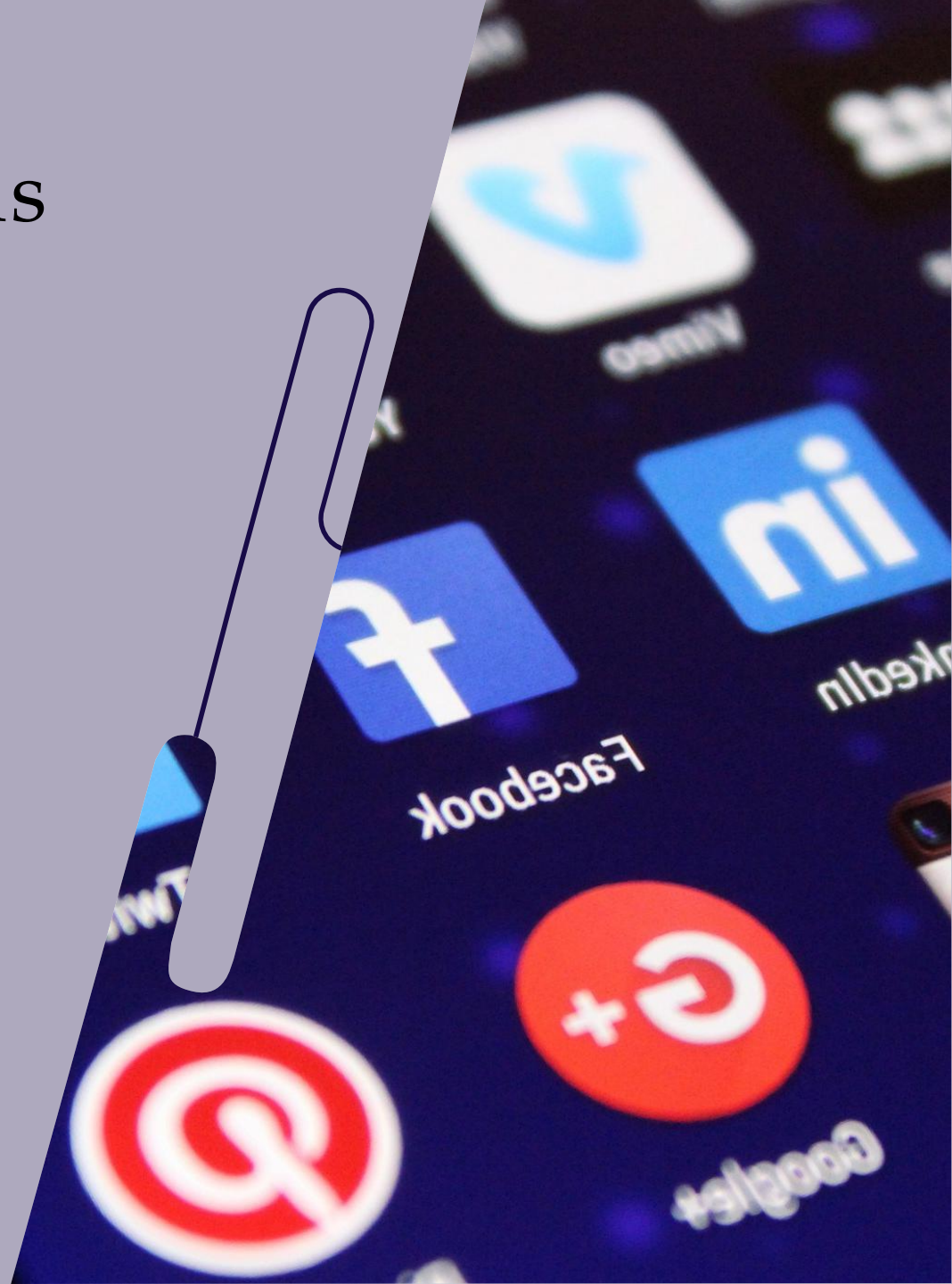
01. How Social Media Spreads Extremist Content

Overview:

- What is Extremist Content?
- How platforms like X amplify it
- Why it spreads so fast

01: How Social Media Spreads Extremist Content

- Extremist content includes material that promotes violence, hatred, or radical ideologies towards specific groups
- X's algorithm prioritizes engagement over accuracy, which pushes extreme content to more users
- Trends and Hashtags can be exploited to push extremist narratives to more people
- Online anonymity lowers social consequences encouraging extremist language

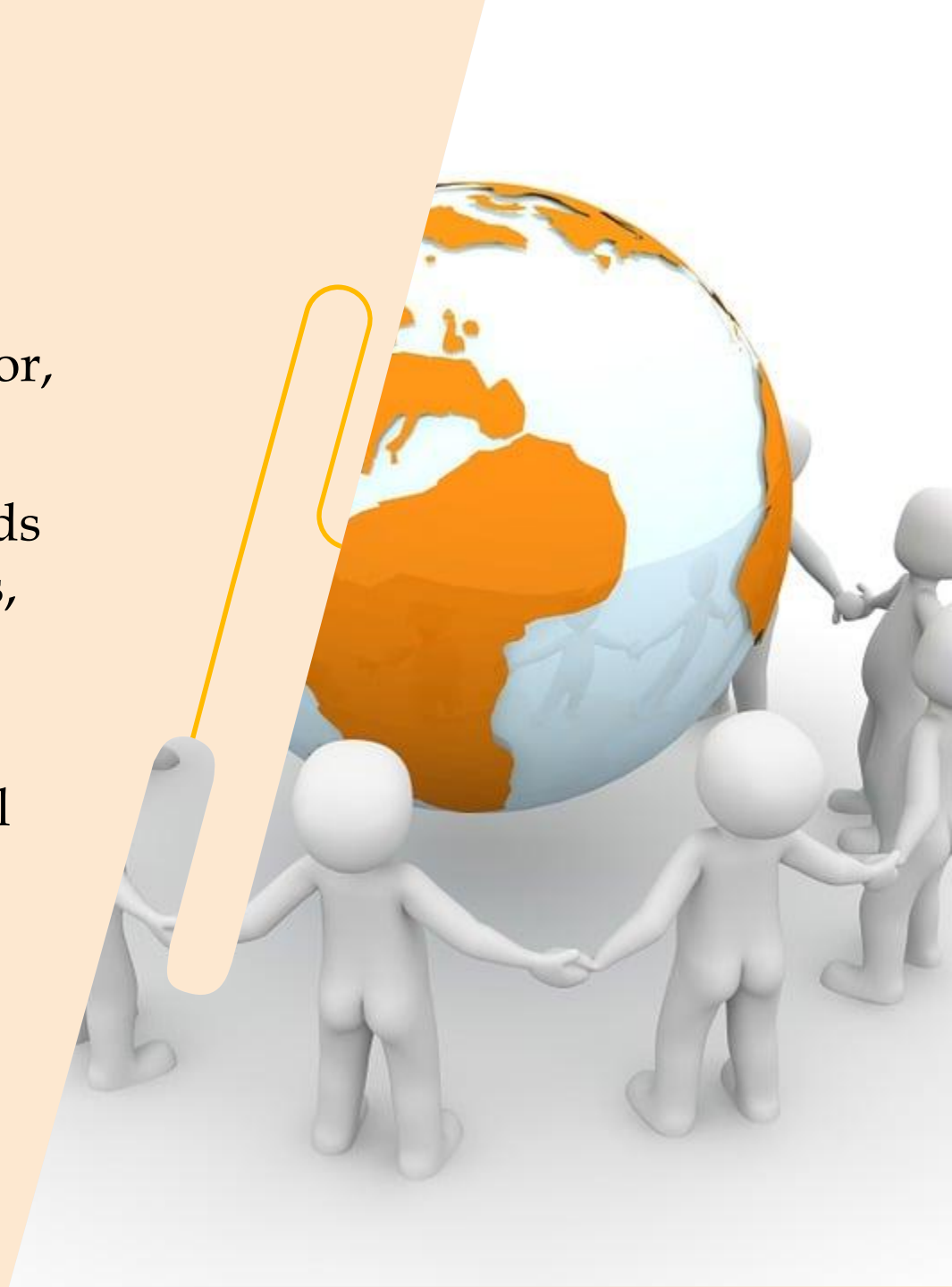


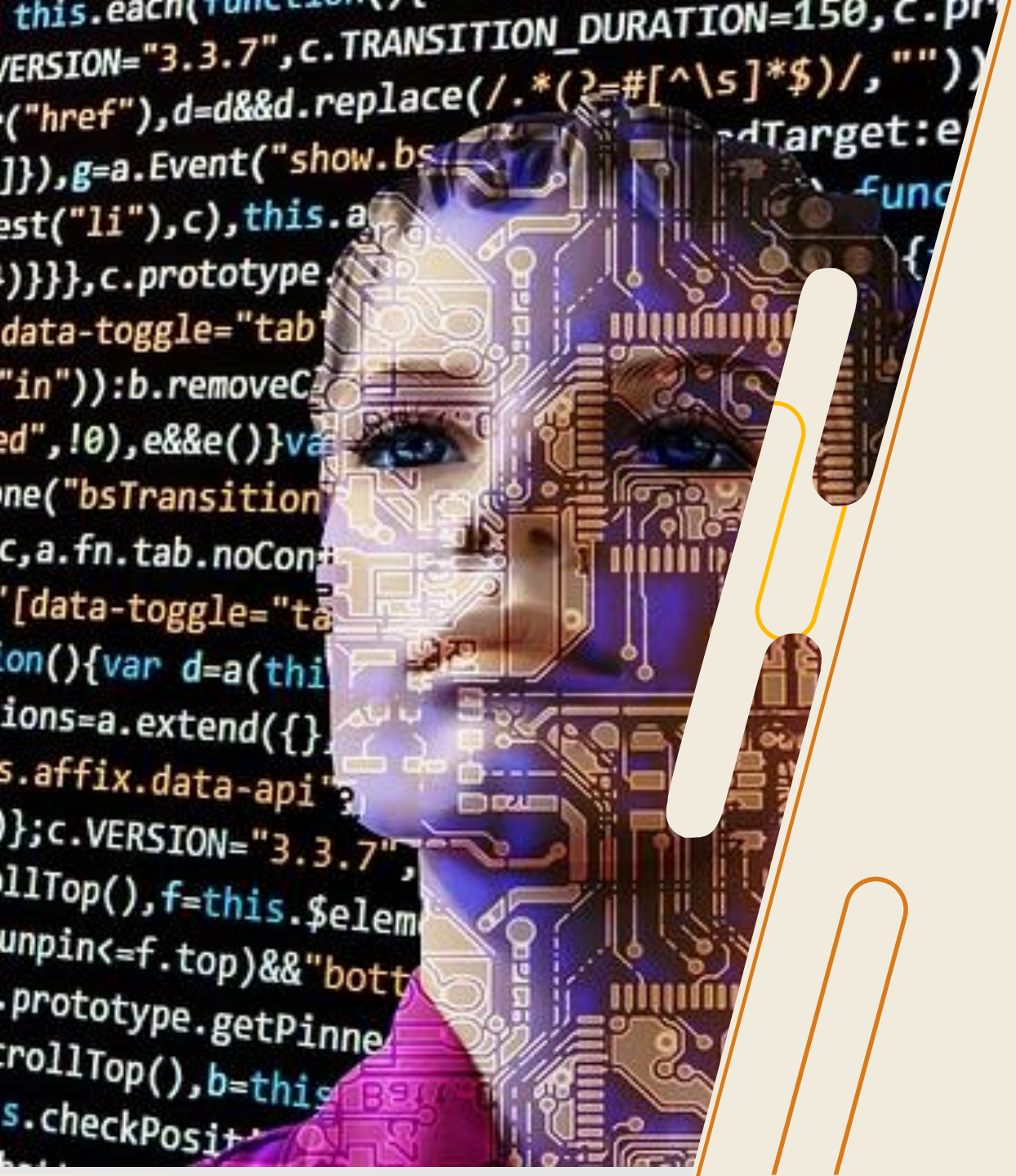
02: The Online Radicalization Process

- Radicalization begins with small, repeated exposures to extremist content that slowly becomes more intense
- X's algorithm accelerates this by continuously serving more extreme content the more a user engages
- Extremist communities use a sense of belonging and identity to pull individuals deeper into radical ideologies
- Users often don't realize they are being radicalized until they are deeply embedded in the community

03: Social Sciences Principles Behind Online Radicalization

- **Relativism –** Changes in technology and social media algorithms directly impact criminal behavior, politics, and social dynamics
- **Confirmation Bias –** X's algorithm constantly feeds users content that aligns with their existing beliefs, deepening extremist views over time
- **Objectivity –** Cybersecurity researchers must study extremist content without injecting personal bias, belief, or opinion



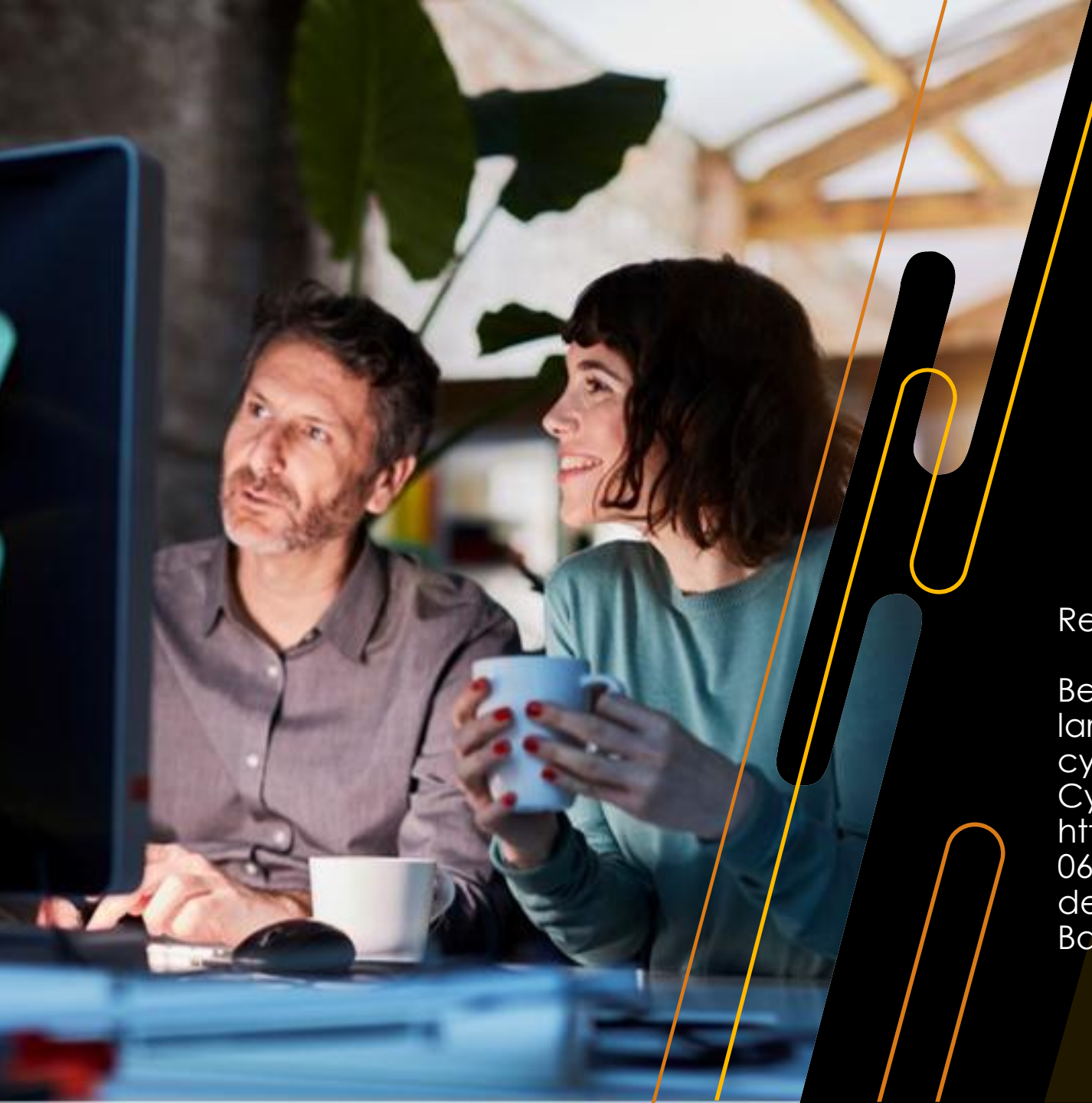


04: LLM-Based Detection Techniques

- **Retrieval-Augmented Generation (RAG) –** Allows LLMs to analyze contextual information from external sources.
- **Linguistic Pattern Recognition –** LLMs analyze tweets for subtle extremist markers, hidden meanings, and geopolitical references
- **OSINT Integration –** Open-Source Intelligence broadens the contextual awareness of LLMs

05: Ethical Challenges

- Marginalized groups are frequently the target of extremist rhetoric
- AI detection models risk false positives
- Researchers and developers must maintain ethical neutrality when studying and moderating extremist content



Thank you

References:

Berzinji, A., & Abdalmajid, M. F. (2024). Utilisation of large language models (LLMs) in OSINT-based cyberterrorism detection on social media. Cybercrimejournal.com. https://www.researchgate.net/publication/389652606_Utilisationof_Large_Language_Models_LLMs_in_OSINT-Based_Cyberterrorism_Detection_on_Social_Media