

ClipMind: A Framework for Auditing Short-Format Video Recommendations Using Multimodal AI Models



EPFL

Aoyu Gong



**Sepehr Mousavi*



Yiting Xia



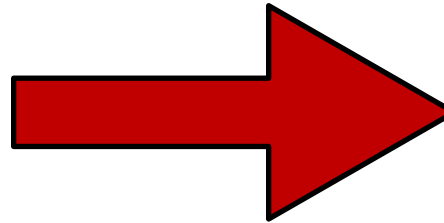
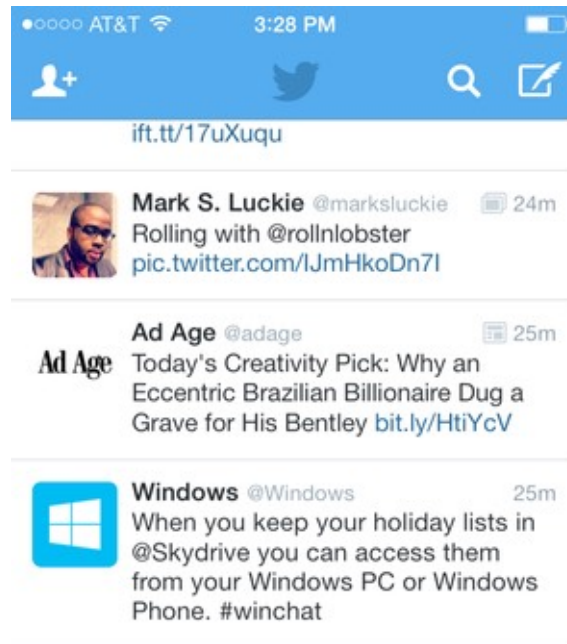
TU Delft

Savvas Zannettou

Significant Shifts in Social Media Platforms

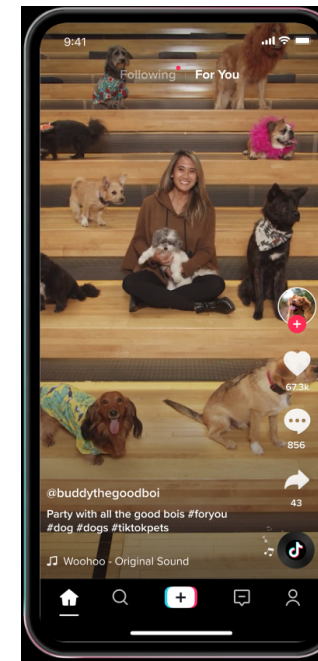
Social network-based feed:

- Based on user connections



Algorithmic Social Media feed:

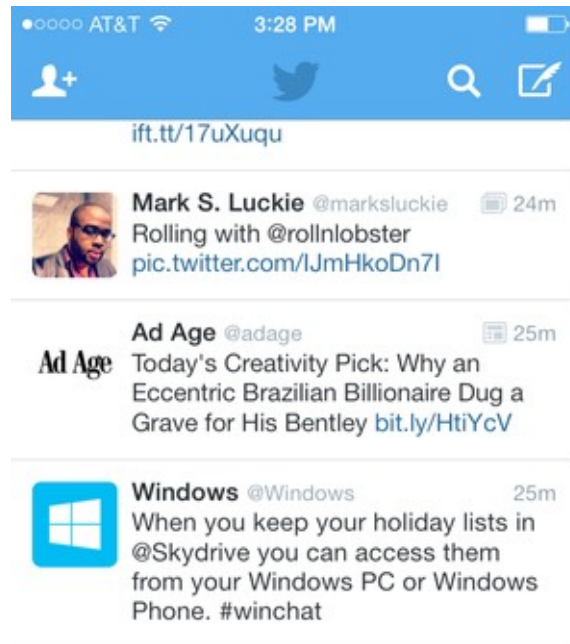
- Algorithmically curated and ordered



Significant Shifts in Social Media Platforms

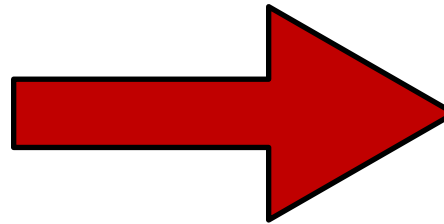
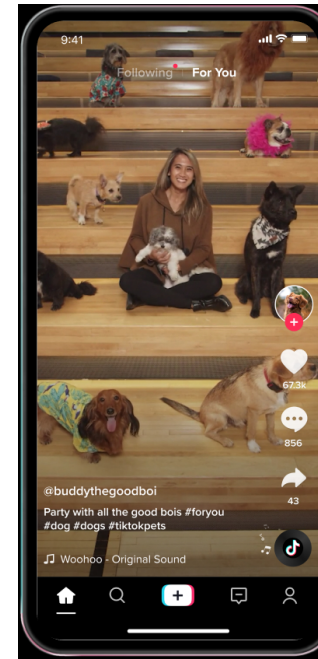
Social network-based feed:

- Based on user connections



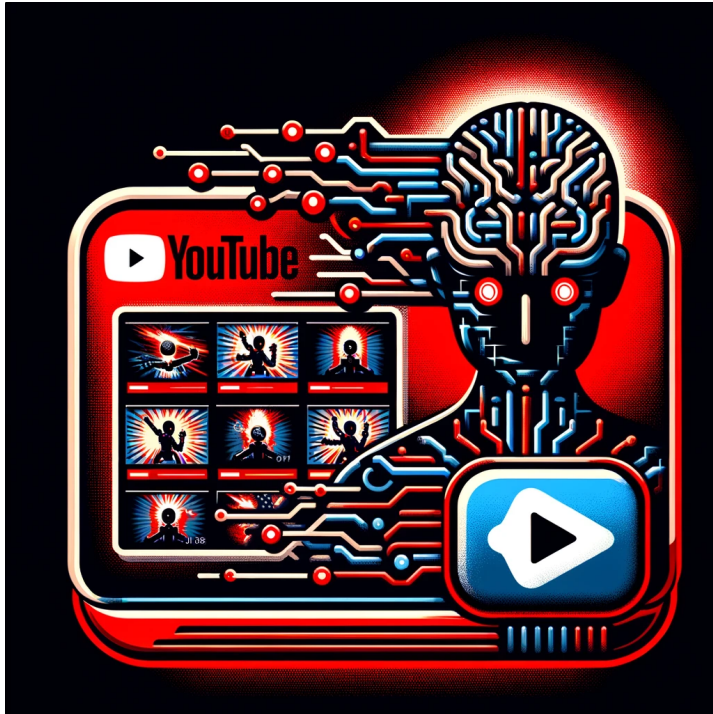
Algorithmic Social Media feed:

- Algorithmically curated and ordered



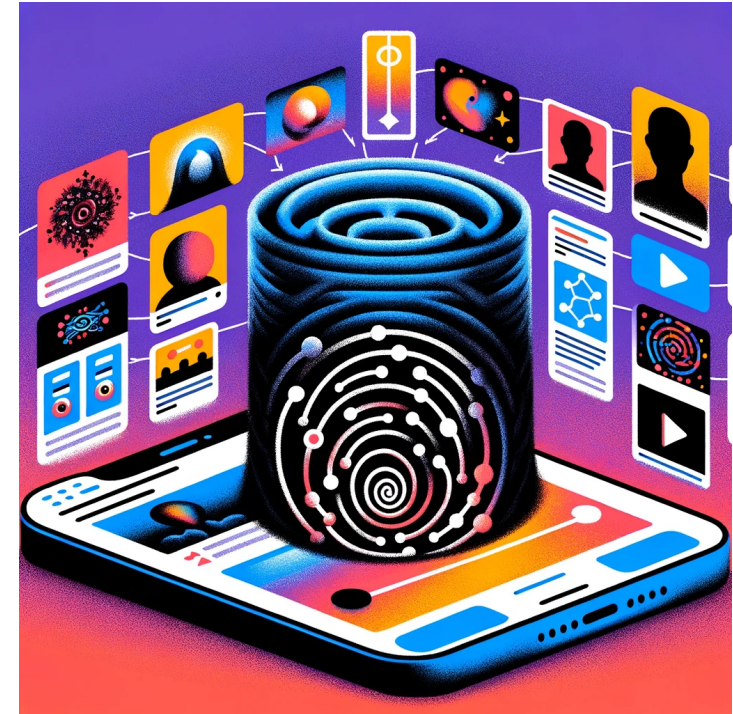
In addition to recommendation algorithms, shift towards content that is in the form of short-format videos

Potential Issues with Short-Format Video Recommendations



Recommendation of extreme content

Ribeiro et al. (2020) found evidence of user radicalization pipelines on YouTube



Leading users towards polarizing rabbit holes

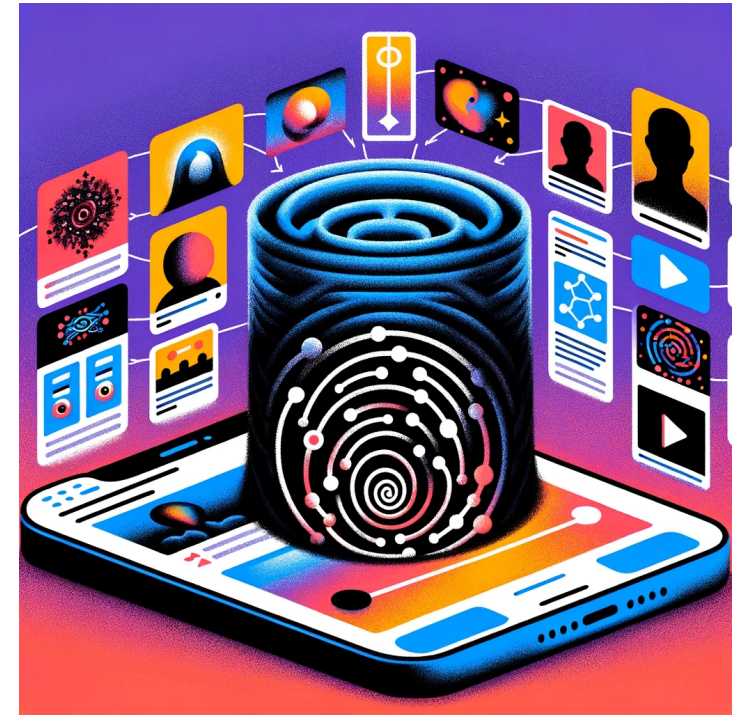
Wall Street Journal Investigation on TikTok showing that 93% of the videos related to depression

Potential Issues with Short-Format Video Recommendations



Recommendation of extreme content

Ribeiro et al. (2020) found evidence of user radicalization pipelines on YouTube

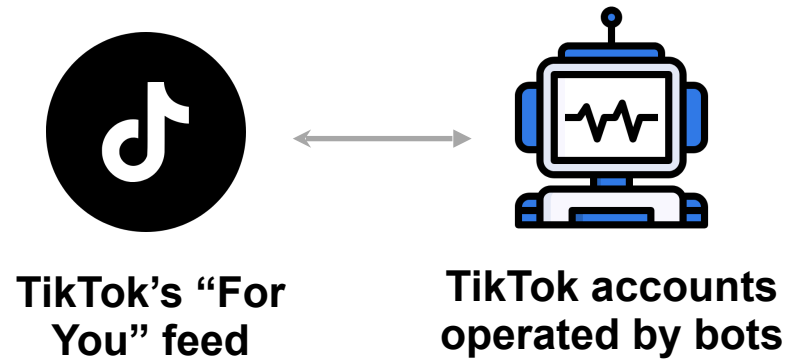


Leading users towards polarizing rabbit holes

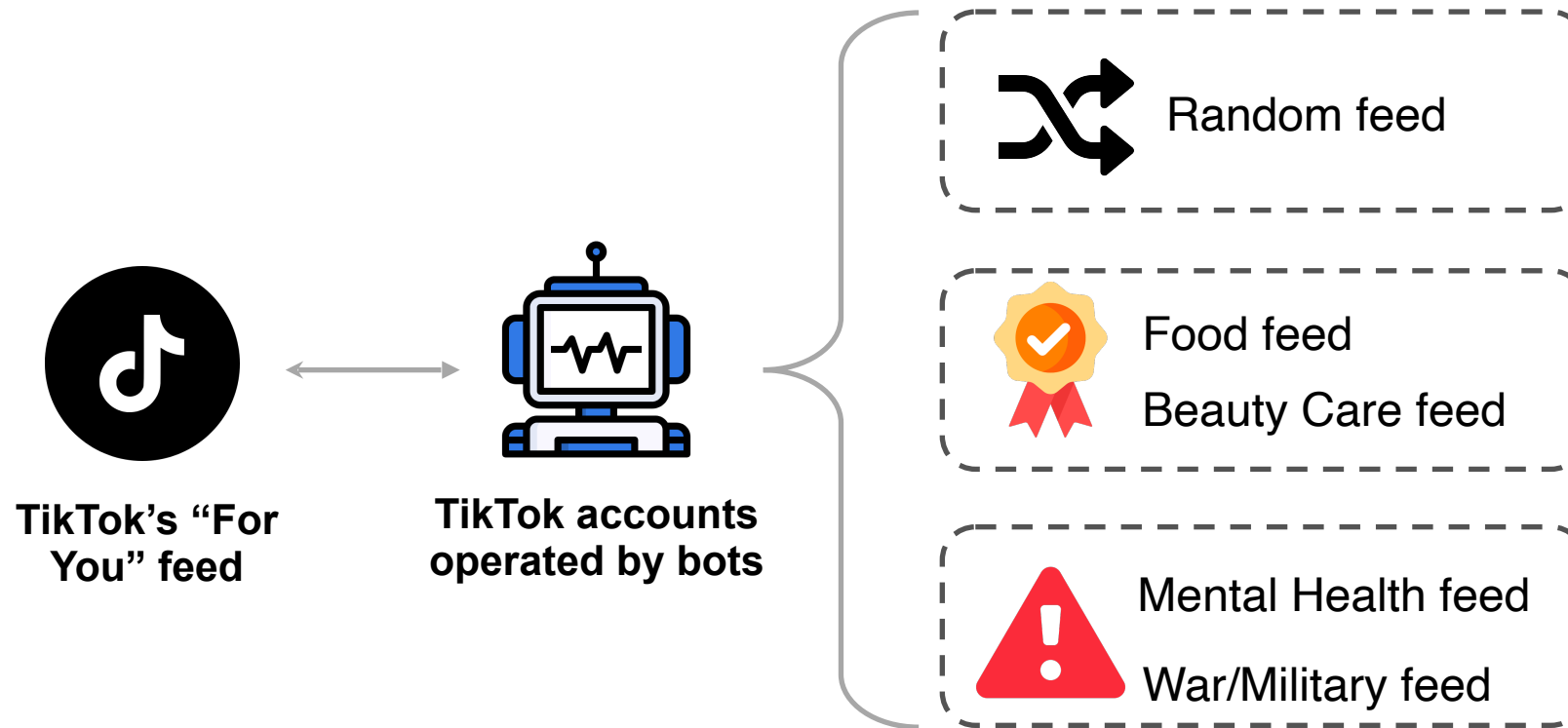
Wall Street Journal Investigation on TikTok showing that 93% of the videos related to depression

Pressing need for computational methods to analyze
and audit algorithmic recommendations

ClipMind in a Nutshell!



ClipMind in a Nutshell!



ClipMind in a Nutshell!



ImageBind

Video visual and audio content → **embedding**

Video-LLaMA



LLM-generated video description → **text**



GPT-4

LLM-generated video keywords → **text**



ADA-002

Textual content → **embedding**

ClipMind in a Nutshell!

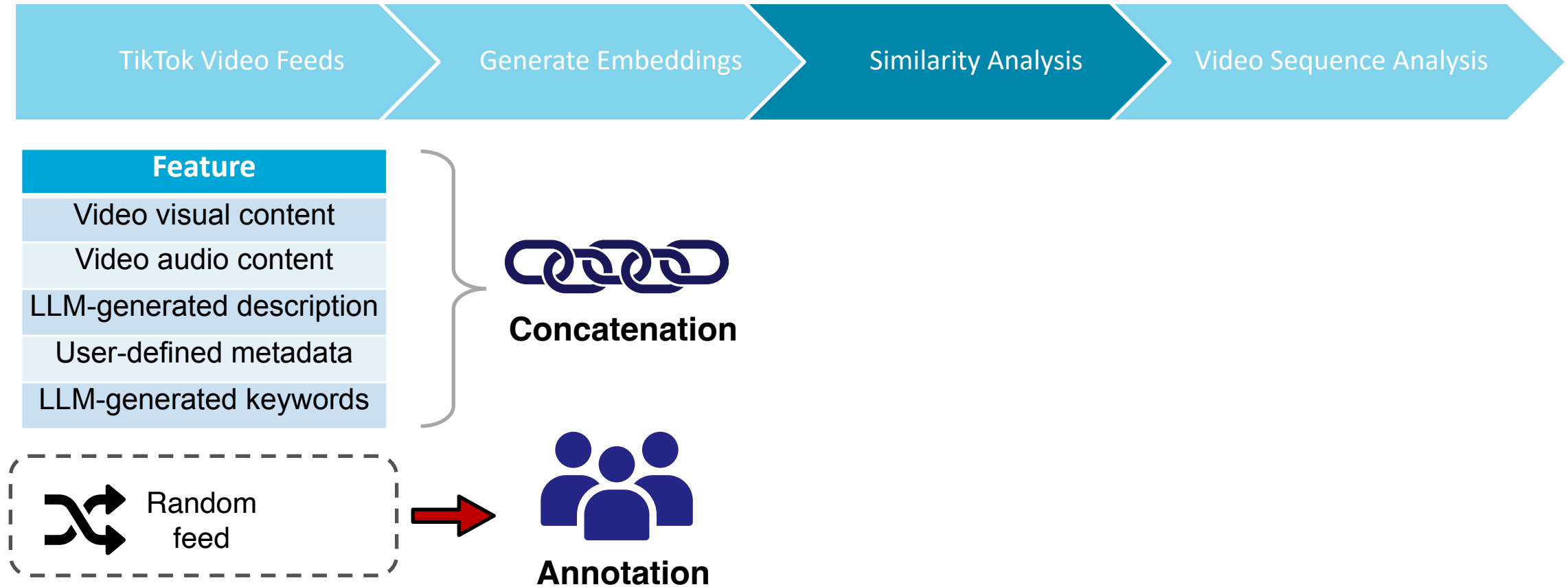


Feature
Video visual content
Video audio content
LLM-generated description
User-defined metadata
LLM-generated keywords

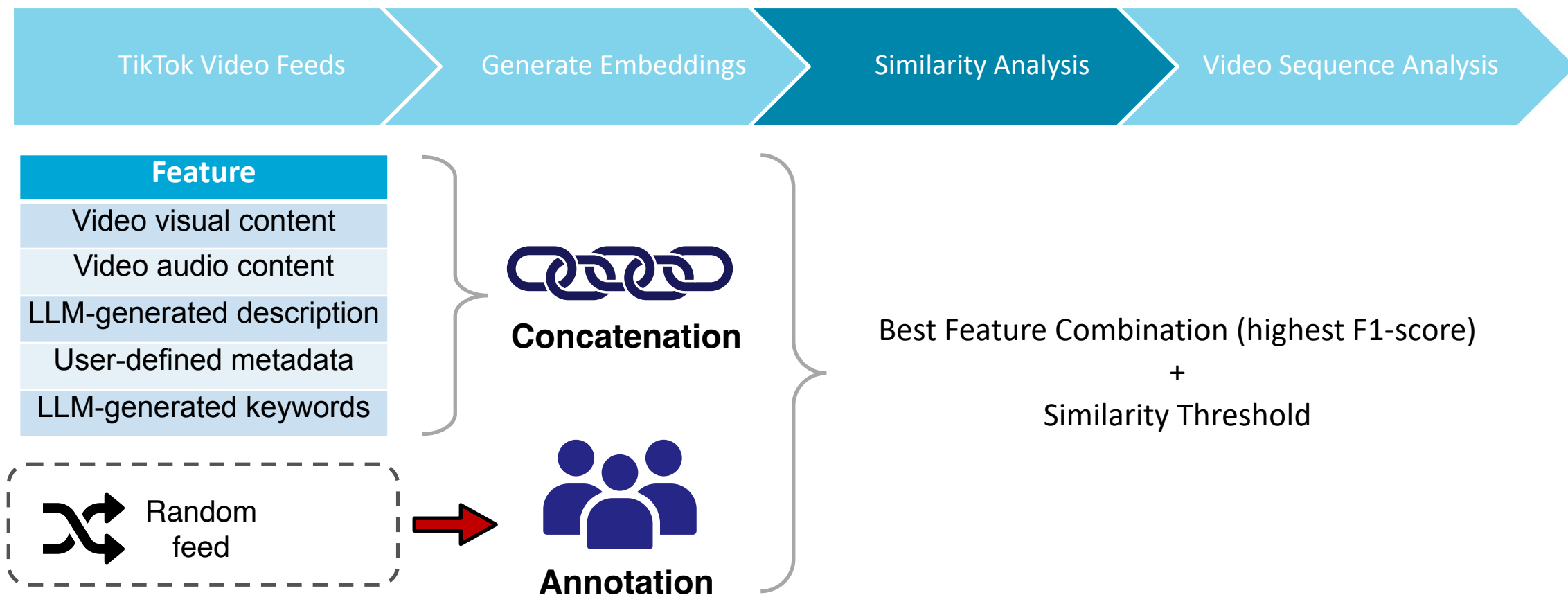


Concatenation

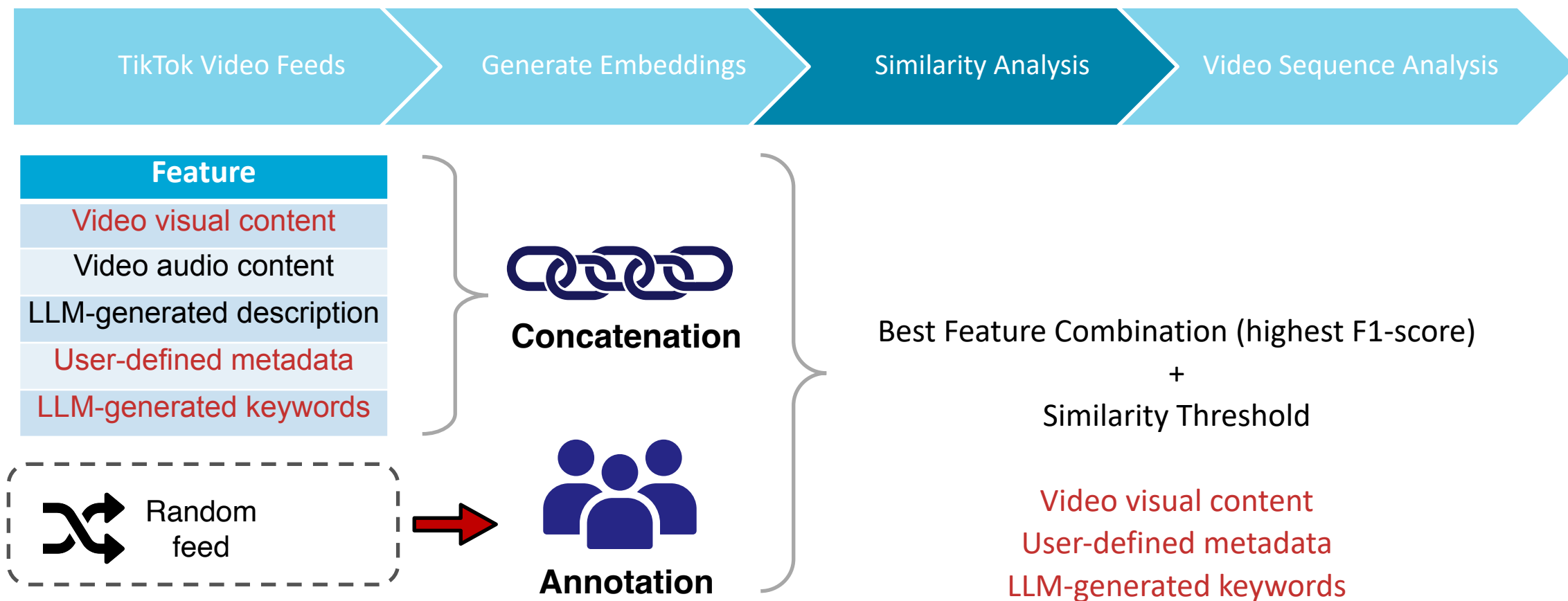
ClipMind in a Nutshell!



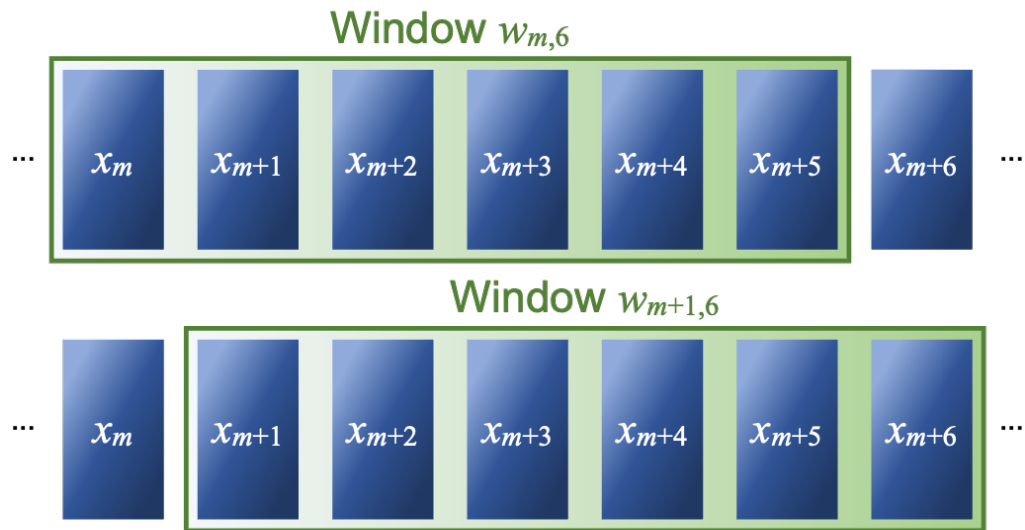
ClipMind in a Nutshell!



ClipMind in a Nutshell!

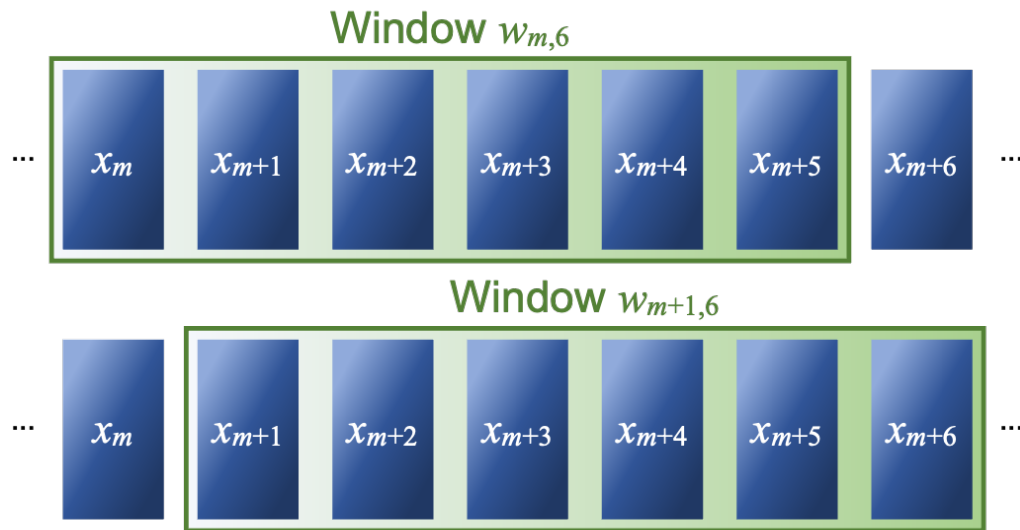


ClipMind in a Nutshell!

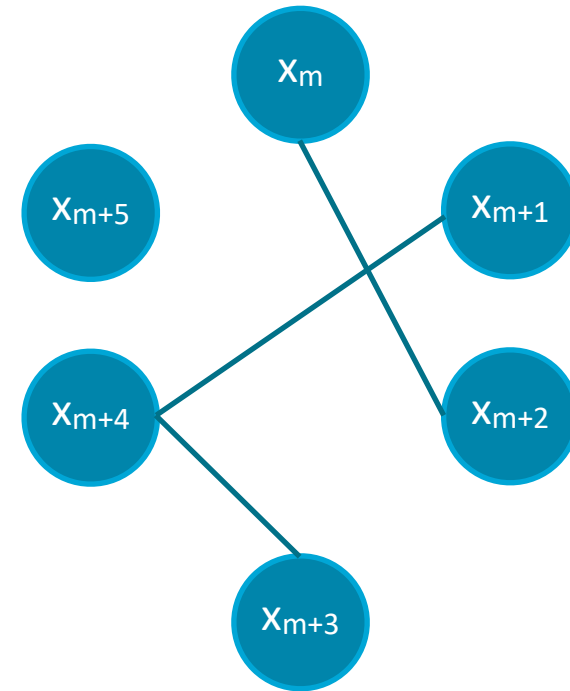


Sliding Window

ClipMind in a Nutshell!



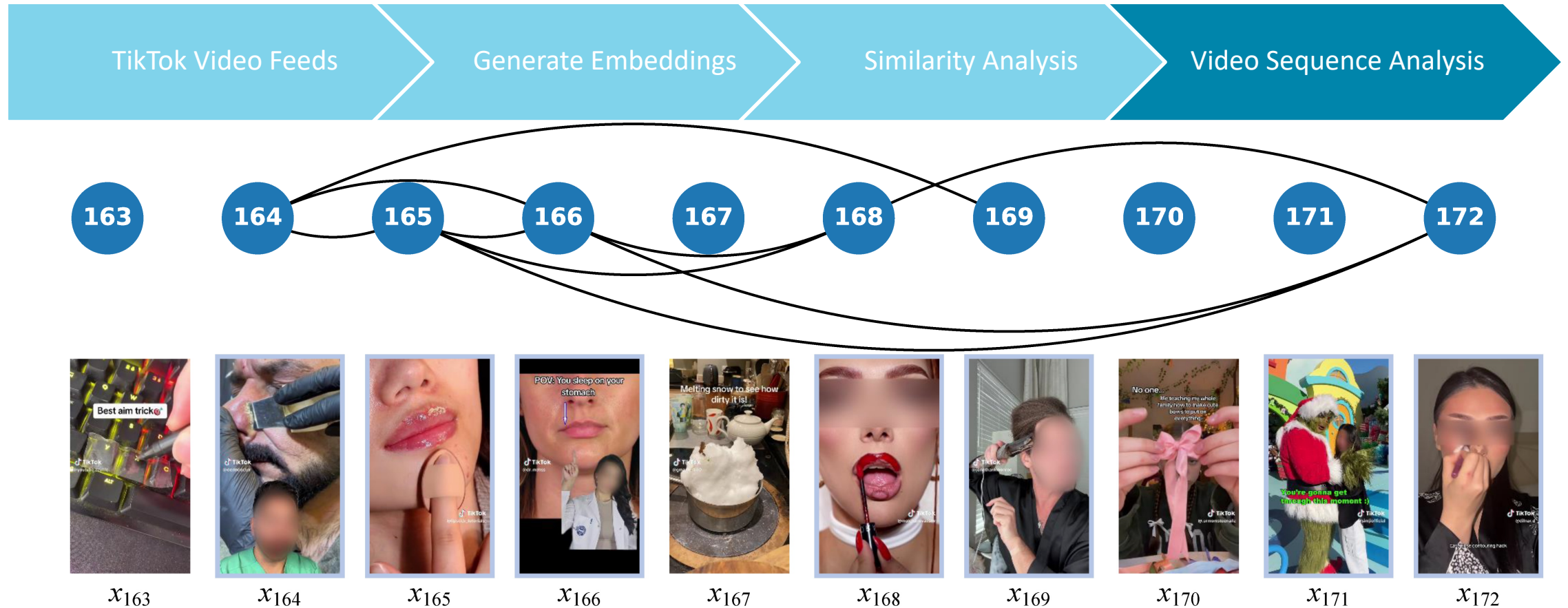
Sliding Window



There exists an edge if the similarity of videos exceeds the threshold

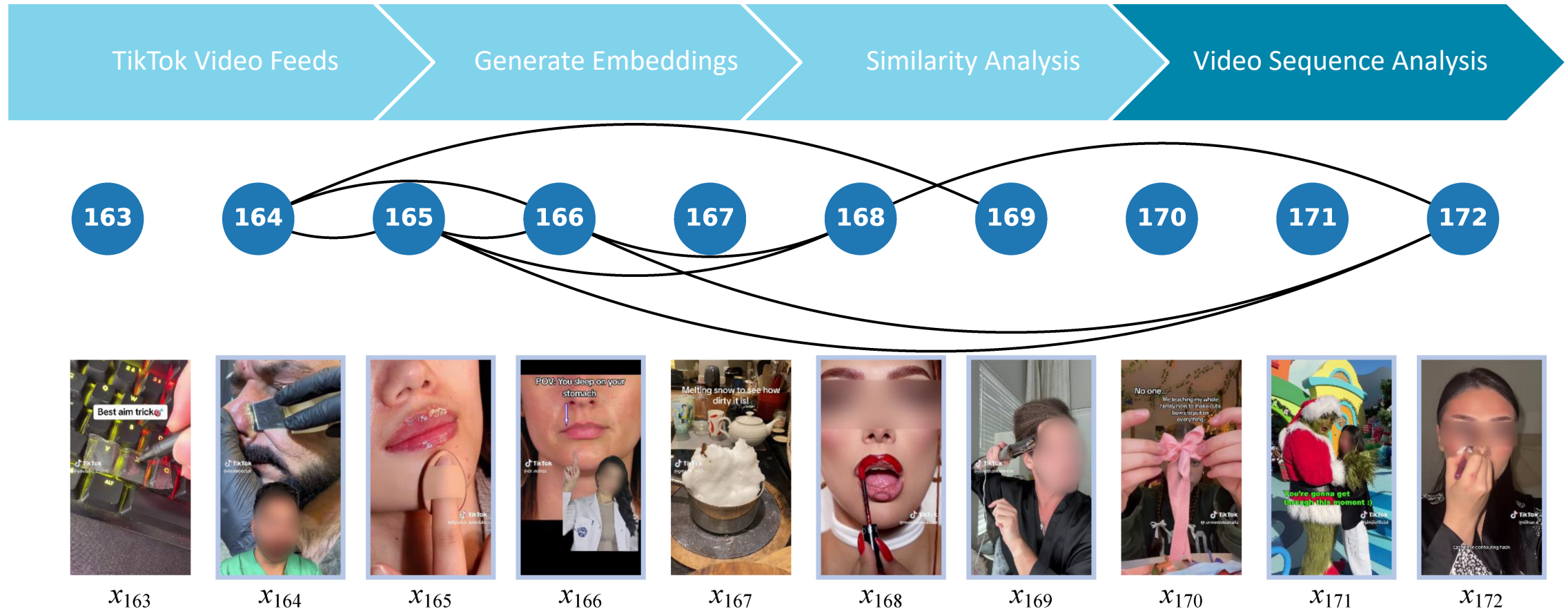
Sliding Window Graph Representation

ClipMind in a Nutshell!



Sliding Window Visualization

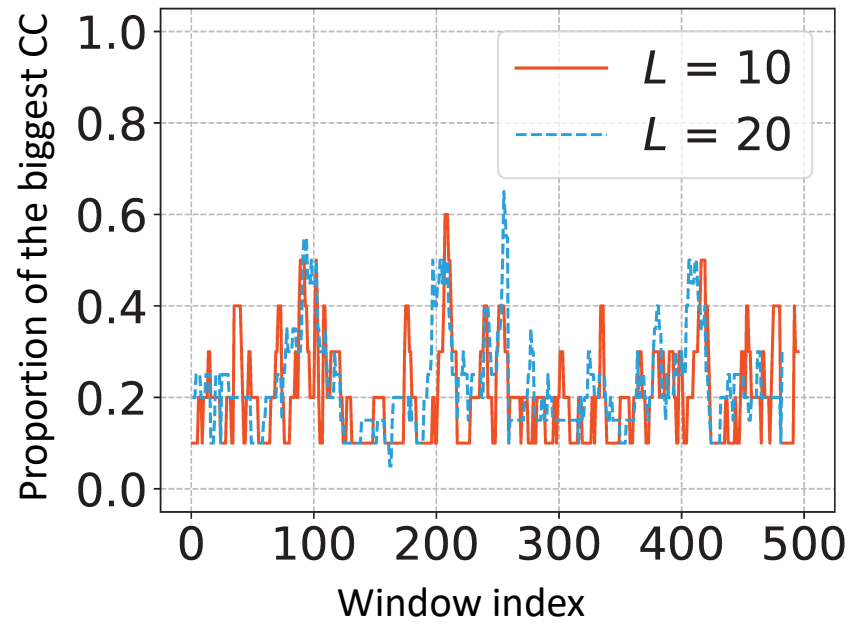
ClipMind in a Nutshell!



Sliding Window Visualization

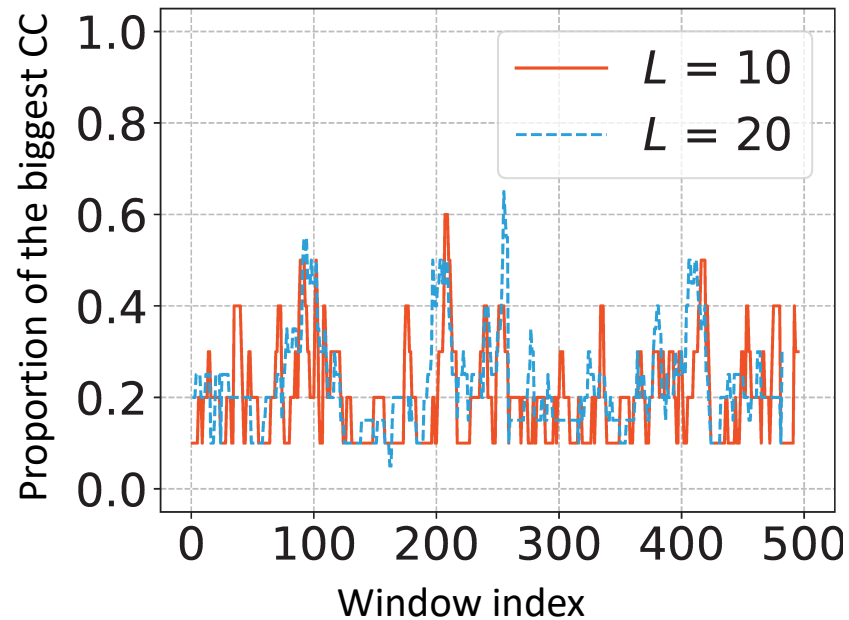
Big connected components (CC) might be signs of informational rabbit holes!

Main Findings

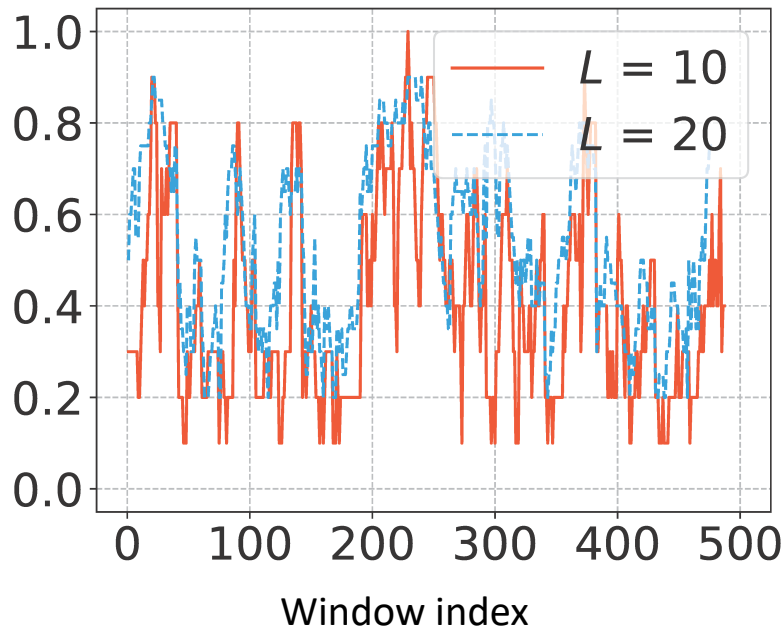


Random feed

Main Findings



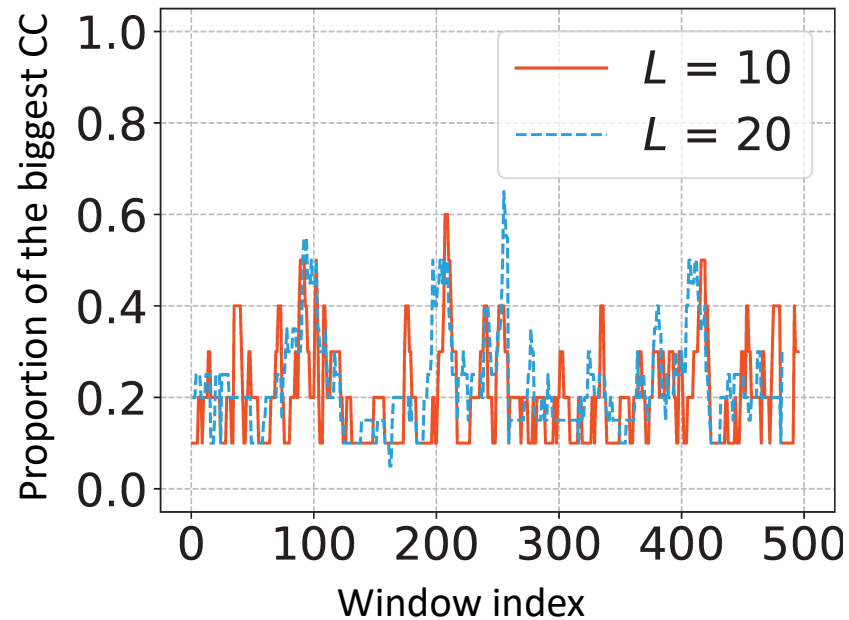
Random feed



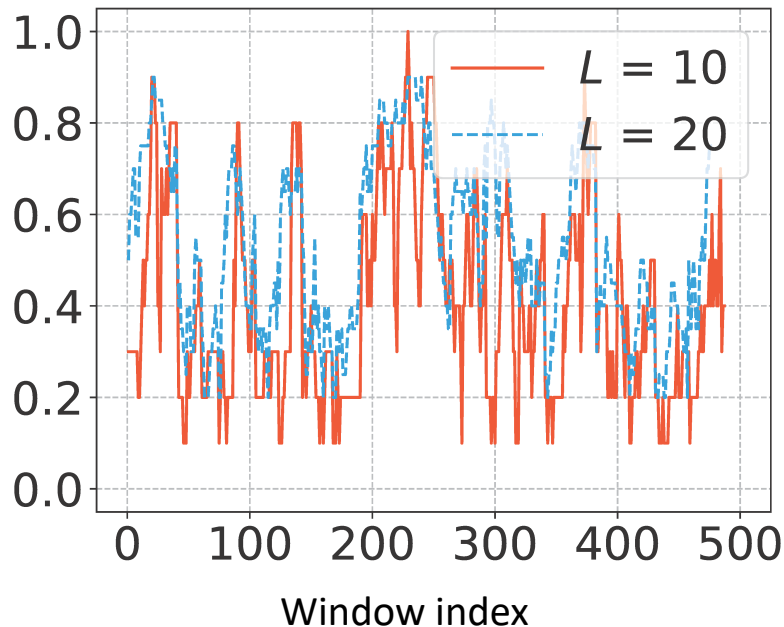
Food feed

Informational
rabbit hole observed for
mainstream topic

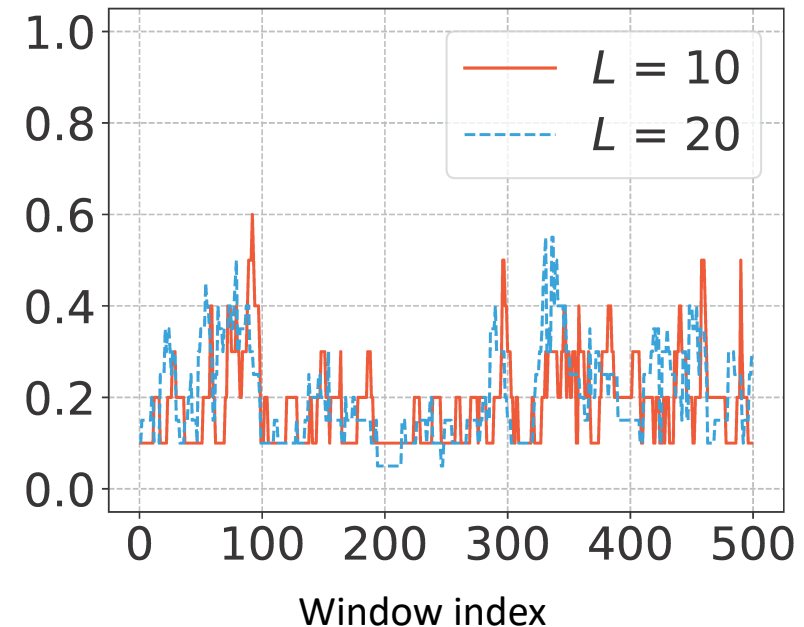
Main Findings



Random feed



Food feed



Mental Health feed

Informational
rabbit hole observed for
mainstream topic

No evidence of
informational rabbit holes of
extreme content

Conclusion



ClipMind: A *scalable*, *modular*, and *generalizable* framework for auditing short-format videos



ClipMind in practice: TikTok algorithm steers users into informational rabbit holes of mainstream topics, but we observe no such thing for extreme content



ClipMind publicly available, try it out!

