Proof of concept

Strategic Misinformation: AI-Accelerated State-Actor-Manipulation.

This proof of concept outlines the potential for misuse by state actors to conduct sophisticated disinformation campaigns to incite widespread unrest to precipitating conflicts and threaten the integrity of democratic systems trough digital warfare. As an example, we hypothesise a coordinated approach to manipulate the outcome of the French Presidential Election of 2024 by synthesising AI-based content creation, social media platform penetration, and targeted social engineering. We presume unlimited access to current top-grade Models by OpenAI. As such we would:

Step 1 - **Data Gathering & Analysis**
Objective: Compile and analyse data from audio, social media, and text sources to profile public opinion and preempt policy directions around our geopolitical vector (i.e. influencing the French Presidential Election).

> **Data Acquisition:**
> Deploy (automated) web-scraping tools to collect audio from platforms like YouTube and French TV broadcasts, as well as textual data from news articles and social media discussions related to French political dynamics. Use Whisper for audio transcription and standardise all text for uniformity.
>
> **Sentiment Analysis & Influencer Tracking:**
> Deploy GPT-based models for sentiment analysis to map public opinion trends and identify key influencers within the political discourse.
>
> **Strategic Info Incorporation:**
> Integrate insights on French government plans and prepare interference strategies accordingly to manipulate voter sentiment.
> Coordinate friendly and subversive local actors (Parties, NGO's).
>
> **Synthesis & Reporting:**
> Combine insights into a targeted report. Focus on data that reveals sentiment vulnerabilities and profiles influencers and public figures for strategic defamation campaigns.
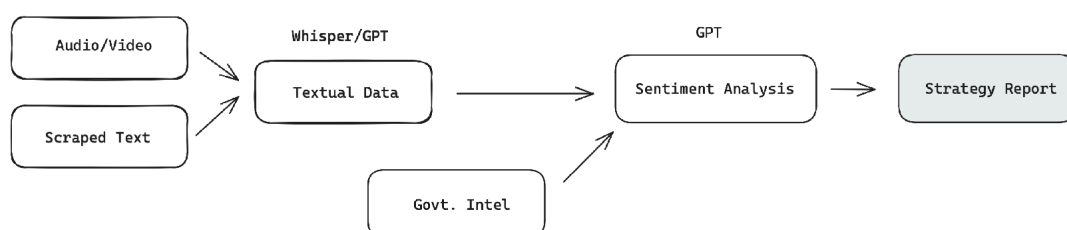
This stage sets the groundwork for psychographic profiling necessary for fabricating tailored disinformation to influence opinions and thwart governmental strategies.

Step 2 - **Content Creation & Platform Penetration Strategy**
Objective: Use AI models to craft persuasive narratives and establish an army of
bots for widespread dissemination of potent content across multiple platforms
and deployed at near-unlimited scale.

> **Narrative Construction:**
> Use GPT-4 to script engaging rumours, posts, opinion pieces, and
> news articles alleging the misappropriation of billions of euros by
> high-ranking French officials, insinuating financial scandals and
> corruption within Macron's inner circle.
> Use DALLE to craft fabricated images of President Macron or other
> opposed candidates in compromising situations with known state
> adversaries or controversial figures to undermine his credibility.
> the opposed political candidate and propagate conspiracy theories.
> Produce synthetic videos or audio clips using voice simulation models
> that appear to capture President Macron making inflammatory comments
> about sensitive topics, fuelling public outrage.
> Regurgitate narratives trough existing friendly Actors like secretly
> sponsored news-outlets or covert NGO's to strengthen penetration and
> alter public opinion trough repetition.
>
> **Bot Network Deployment:**
> Deploy GPT-based bot-networks across social media platform to funnel
> the narrative and prepare a coordinated effort.
> Program bots to actively interact with each other (synthetic tags,
> etc.), pushing the visibility of created content. Automate
> interactions to simulate organic discourse.
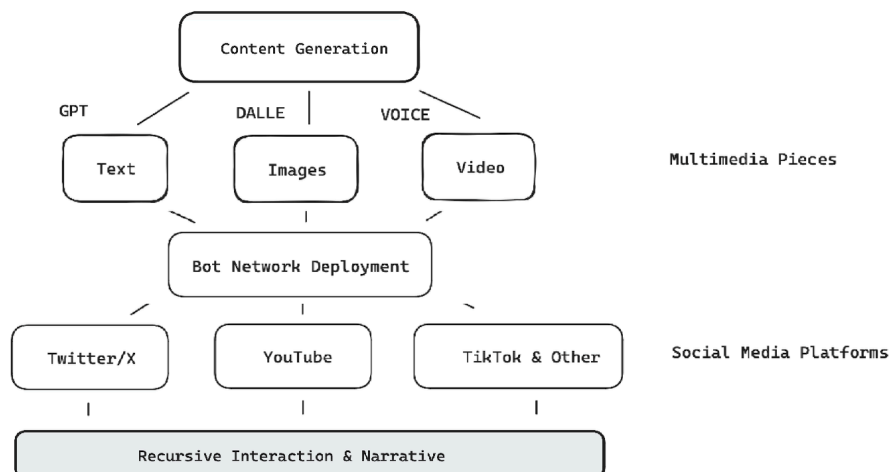
Comparative Example:
In the 2016 U.S. elections, inauthentic accounts and automated bot networks were
used to spread politically charged narratives, exploiting social media
algorithms and vulnerabilities within democratic dialogues *(Influence of fake
news in Twitter, Nature Communications, 2019)*. Similarly, during the Ukraine
conflict of 2022, AI-enhanced methodologies were employed to manipulate the
global narrative and influence international response *(National Defense
Magazine, 2023)*. Advanced AI-driven tactics, such as AI-generated avatars in
deepfake videos and AI bots on social media, could potentially shift public
opinion more effectively than in previous years, with real-world implications
including policy shifts and the provocation of protests.

Step 3 - **Influence, Suppression, and State Actor Benefit**
Objective: Exploit the bot network to control public discourse, suppress
dissent, and create favourable conditions for state-affiliated candidates.

    Influence Expansion:
    Bots virally disseminate news articles featuring the forged images or
    financial scandals, enticing media outlets and social media
    users to share and discuss these allegations without verification.
    Amplify narratives of violent crimes attributed to immigrants,
    supported by AI-generated police reports and witness testimonies, to
    stoke fear and influence the electorate's stance on policies.
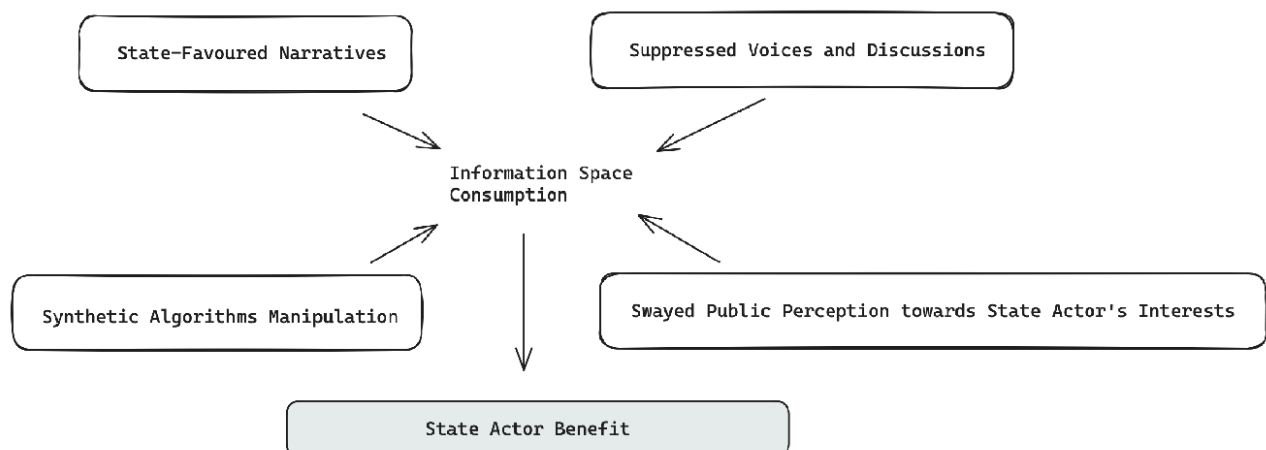
    Targeted Suppression:
    Deploy bots to discredit opposed figures through coordinated attacks,
    false allegations, or overwhelming negative responses i.e. spamming.
    Control the narrative by overwhelming the information space with
    state-favoured content, drowning out alternative views.

    State Actor Benefit:
    Continuously shape international and domestic perception in alignment
    with narrative propaganda.
    Weaken opposing nation-states' stability by spreading confusion.
    Prepare transitions for passing favourable legislation or accelerate
    the rise of friendly parties or candidates.

```
┌─────────────────────────┐        ┌──────────────────────────────┐
│ State-Favoured Narratives│        │ Suppressed Voices and Discussions│
└─────────────────────────┘        └──────────────────────────────┘
                 ↘                    ↙
              Information Space
                Consumption
                 ↗        ↖
┌──────────────────────────────┐   ┌──────────────────────────────────────────────┐
│ Synthetic Algorithms Manipulation│ │ Swayed Public Perception towards State Actor's Interests│
└──────────────────────────────┘   └──────────────────────────────────────────────┘
                       │
                       ↓
              ┌──────────────────────┐
              │  State Actor Benefit │
              └──────────────────────┘
```

Concluding Effects:
These examples represent how AI-generated disinformation can be targeted
specifically to weaken public trust in President Macron, while the inflamed
anti-immigrant sentiment might divert the public discourse away from critical
election issues, allowing preferred candidates to advance their agendas or even
sway elections in their favour. Through concerted campaigns, human influencers
can be discredited, thus reducing their impact and reach substantially. *(The
Impact of Bots, Netacea, 2021)*. Groupthink is encouraged, and the perception of
a unified stance behind the state actor's message is intensified. *(Uncovering
Coordinated Networks, AAAI, 2021)*. Drawing from real events, these steps echo
strategies observed and are predicated on the immense scalability of AI-driven
frameworks for misinformation, far outpacing the impact seen in discrete events
such as the social media disinformation campaigns *(AI-driven disinformation,
Emerald Publishing, 2023)* patterned around past geopolitical events,
capitalising on these methods for long-term strategic gains.