# Additional Application Material

Chi Shun Fong
Ph.D in Political Science and Social Data Analytics
The Pennsylvania State University

## Table of Contents

| | |
|---|---|
| 3. Research Statement | P.2 – P.3 |
| 4. Teaching Statement | P.4 – P.5 |
| 5. Transcript (PhD) | P.6 – P.9 |
| 6. Departmental Certification | P.10 |
| 7. Sample Syllabus – Computational Social Science | P.11 – P.15 |
| 8. Sample Syllabus – Politics and Development of Contemporary China | P.16 – P.20 |
| 9. Writing Sample 1: Western Influencers as Chinese Government Foreign Defenders | P.21 – P.77 |
| 10. Writing Sample 2: A Replicable Visual LLM Workflow for Emotion Coding in Political Events | P.78 – P.128 |

## Research Statement

Situated at the intersection of authoritarian politics, contentious politics, political communication, and computational social science, my research examines how state and non-state actors—from official media to influencers— manipulate public opinion and information environments to maintain authoritarian stability in today's fragmented digital ecosystem. My research is organized across three interrelated streams: **(1) influencer-assisted transnational propaganda, (2) adaptive information control and domestic contention, and (3) computational measurement of latent public opinion**

**Influencer-assisted transnational propaganda:** Authoritarian regimes actively seek support from foreign audiences to foster economic development and mitigate anti-regime pressure. However, direct state-sponsored foreign propaganda often triggers audience skepticism. I argue that authoritarian regimes circumvent this credibility gap by co-opting seemingly autonomous influencers who possess local identity and cultural resonance. In my first dissertation chapter, I conduct a survey experiment demonstrating that U.S. citizens distrust videos endorsed by Chinese state media on human rights in China, but not those by white influencers. This trust deficit is most pronounced among the targeted subpopulation with low media trust and negative attitudes toward the Chinese government. Expanding on these findings, my ongoing digital fieldwork traces how pro-China influencers' activities are expanding to developing regions along China's Belt and Road Initiative to project the soft power of the Chinese government.

**Authoritarian information strategies and domestic contention:** Information control is critical to authoritarian resilience, but blunt censorship or repression risk backfiring in this digital era. My studies show how authoritarian regimes avoid collective action through adaptive and preemptive information strategies. My studies found that the Chinese government handles grievances over the nationwide expansion of environmental projects by promoting post-implementation solutions, and a follow-up experiment found that these seemingly responsive, yet highly controlled solutions could effectively demobilize citizens with low trust in government and prior exposure to environmental activism. Complementing this, my second dissertation chapter helps explain why the Chinese government is increasingly co-opting or repressing online opinion leaders. Using a survey experiment, I found that these leaders mobilize Chinese citizens to participate in environmental activism by increasing their expectation of mass participation, which signals a lower risk of repression and a higher likelihood of success.

**Computational measurement of public opinion:** I develop low-cost, scalable, and accurate computational tools to study increasingly complex and multimodal online data. In my third dissertation chapter, I construct a workflow that extracts emotions, a critical mobilizing resource, in social media images with near-human accuracy using LLMs that could code thousands of images within one day for less than $10 USD per thousand images. Applying this workflow to study social media data related to the 2019 Hong Kong protests, I demonstrate that visual expressions of solidarity attenuate the demobilizing effects of state repression. Beyond visual analysis, I design and apply methods to uncover latent attitudes to study public opinion, including deep learning models to assess US subnational political elites' stances on the integrity of the 2020 election based on their tweets and double-list experiments to measure regime support in Iran amid preference falsification.

**Future research plan**
I plan to extend the scope and regional focus of my work on authoritarian propaganda and further advance the related computational measurement of public opinion.

**Everyday governance and performance legitimacy**: I will expand my foreign propaganda framework from high-level ideological narratives to everyday governance issues, such as personal safety, the cost of living, government efficiency, and public infrastructure. These narratives contrast authoritarian state capacity with perceived democratic dysfunction. Through survey experiments across democracies and autocracies, I will test whether exposure to these narratives shifts respondents' support toward authoritarian-style governance and norms, as well as support for their home government. I will also explore the psychological mechanisms that drive susceptibility to these narratives among specific subpopulations. This initiative would also contribute to our understanding of both authoritarian stability and democratic backsliding.

**Chinese Propaganda in Asia**: Focusing on East Asian and Southeast Asian countries, which are more culturally similar to China than to Western countries, I will use content and network analyses to examine how Chinese state-linked actors tailor influencer collaborations and cultural alignment to achieve local appeal in Asia, especially those involved in China's Belt and Road Initiative. This project maps how China adjusts its external persuasion strategies as it shifts from adversarial Western platforms and audiences to deeply interconnected regional neighbors.

**Confirmation bias in AI-generated content**: I argue that political orientation shapes judgments of authenticity: people perceive AI-generated messages aligned with their beliefs as more "human-like" and credible. Through field experiments across countries, I will test how belief congruence, regime type, and media literacy affect trust in AI-generated content and evaluate whether AI-generated personas can replicate the authenticity of online opinion leaders.

**Measurement of Multimodal Public Opinion**: I will move beyond text and images to include video and audio in my LLM-based multimodal analysis. This move is particularly helpful, as social media data has become increasingly multimodal and requires analysis that involves not only different data modalities but also integrated analysis. This will support research beyond my substantive research on propaganda to other subfields and topics of interest to social scientists, such as electoral campaigning, misinformation, and parliamentary communication.

`

## Teaching Statement

### Introduction

Teaching, for me, is a collaborative process that bridges my research expertise with student learning, which I will outline step by step here. First, I outline my **pedagogical approach,** which encourages students to learn through active engagement, critical inquiry, and collaboration with peers, while also providing them with up-to-date knowledge in the field under my instruction. The following is my **commitment to diversity,** which aims to address the unique challenges faced by students from diverse backgrounds. Next, I discuss my distinct sets of **learning goals for undergraduate and graduate students**, aimed at helping them pursue both non-academic and academic careers. Finally, I conclude with my **teaching interests and experience**.

### Pedagogical Approach

I believe in interactive, inclusive, and student-centered teaching. Students, especially graduate students, are collaborators, not passive recipients. I encourage their active participation through in-class discussions, group work, and tutorials while providing constant guidance and feedback. In addition to my teaching experience, I also draw inspiration from the training provided by the **Schreyer Institute for Teaching Excellence** at Pennsylvania State University (PSU). Here are some of my past practices.

- *Personalized Course Design*: At the start of each semester, I conduct an anonymous survey to understand students' backgrounds, expectations, and prior knowledge, allowing me to adjust the course content accordingly. For instance, I will include more articles as reading material rather than textbooks when students have a stronger foundation
- *Recap and Connection*: I begin each class with a five-minute recap of the previous concepts, followed by a news headline or current event that connects to the class material.
- *Active Learning*: In my "China in the Era of Globalization" class, I split students into groups to debate the effectiveness of China's anti-corruption campaigns under Xi Jinping's reign, requiring each group to support its arguments with empirical data.
- *Writing Workshops*: To help students develop their writing skills, I organize workshops in which they learn from published papers and receive feedback on their drafts.
- *Real-World Application*: In my "Introduction to Social Data Analytics for Public Affairs" course, I asked students to extract data from YouTube to analyze differences in the Chinese government's narratives on the Russian-Ukrainian conflict across its domestic and international media outlets.

### Commitment to Diversity

I am deeply attentive to the diversity of student backgrounds and learning needs. My experience teaching at public universities and community colleges has taught me the importance of inclusivity.

- *Adaptive teaching*: I incorporate audiovisual materials, such as movies and music, to illustrate complex concepts for students without a strong prior background in social science. I once played the film *Thirteen Days*, which is about the Cuban Missile Crisis, to introduce game theory to students.
- *Language accommodation*: As I pursued my bachelor's degree and taught at the Chinese University of Hong Kong, where Chinese is also a medium of instruction, I supplemented English instruction with Chinese for students who struggled with English.

- *Hands-On Learning*: In methods classes, I incorporate code visualization, interactive simulations, and hands-on coding sessions to support students challenged by math.
- *Flexible Support*: I have worked closely with students from the military and law enforcement, as well as those facing mental health challenges, by offering flexible deadlines, individualized feedback, and out-of-class appointments.
- *Alternative Assessments:* In the course "Media in the Age of Globalization," I allowed students to choose among traditional essays, podcasts, and infographics for their final projects, accommodating diverse strengths and learning styles.

**Learning Goals for Undergraduate and Graduate Students**

I set different expectations for students. In undergraduate courses, I focus on introducing students to a social science perspective, using real-world events to inspire their interest in learning. In this regard, I assess the extent to which they can illustrate and present facts using learned concepts in quizzes, discussions, and papers, using existing datasets or simple data-collection methods. It helps students develop their perspective for future studies and polish their transferable skills, such as data summary and presentation. In graduate courses, I emphasize theoretical engagement with empirical evidence, so that graduate students must critically evaluate existing knowledge about the data-generating process and the methodology, thereby developing their scholarly voices. I will share my research experience, findings, and cutting-edge techniques I acquired with graduate students. I ask graduate students to develop an executable proposal for future publication in substantive courses and complete a data replication with improved or alternative methods in methodology courses.

**Teaching Interests and Experience**

My teaching interests closely align with my research in authoritarian politics, political communication, comparative politics, and quantitative methods, allowing me to provide the most up-to-date knowledge in the field and to discuss the challenges I face in my research process in class. I am happy to teach the following courses (* with syllabus developed):

- **Substantive**: Introduction to Comparative Politics*, Authoritarian Politics, Social Movement and Contentious Politics*, and Media and Politics*
- **Methodological**: Quantitative Methods*, Research Design, Machine Learning/Text as data with Python, and Social Data Analytics*
- **Regional**: Politics and Development of China*, and Politics of Hong Kong

Here are some courses that I have taught

- **Substantive**: Introduction to Comparative Politics (PSU, Summer 2025)
- **Methodological**: Introduction to Social Data Analytics for Public Affairs (The Hang Seng University, Fall 2023 - 100% satisfaction)
- **Regional**: Government and Politics of Hong Kong (Open University of Hong Kong, 2017, 2019 – 98% satisfaction)

Additionally, I have served as a teaching assistant in Hong Kong and the US for a wide range of courses (see my CV) and received the Excellent Teaching Assistant Award from the Chinese University of Hong Kong in 2018.

# Graduate Official Transcript

**Name:** **Chi Shun Fong**
**Student ID:** **940461840**
**SSN:** *****-**-8313**
**Birthdate (MM/DD):** **06/04**

Print Date: 08/04/2025

mady forsythe

Degrees Awarded

Degree: Master of Arts
Confer Date: 12/17/2022
Plan: Political Science (MA)

Beginning of Graduate Record

Fall 2020

Program: Liberal Arts
Plan: Political Science (PHD) Major

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| PLSC | 501 | Meth Pol Analysis | 3.000 | 3.000 | A | 12.000 |
| PLSC | 502 | Stat Meth Pol Res | 3.000 | 3.000 | A | 12.000 |
| PLSC | 560 | Intnl Rel Thry/Mth | 3.000 | 3.000 | A | 12.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 4.000 | Term Totals | 9.000 | 9.000 | 9.000 | 36.000 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 4.000 | Comb Totals | 9.000 | 9.000 | 9.000 | 36.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 4.000 | Cum Totals | 9.000 | 9.000 | 9.000 | 36.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 4.000 | Comb Totals | 9.000 | 9.000 | 9.000 | 36.000 |

Spring 2021

Program: Liberal Arts
Plan: Political Science (PHD) Major

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| PLSC | 503 | Multi Anly Pol Res | 3.000 | 3.000 | A | 12.000 |
| PLSC | 511 | Professional Norms | 1.500 | 1.500 | P | 0.000 |
| PLSC | 550 | Com Pol Thry/Meth | 0.000 | 0.000 | LD | 0.000 |
| PLSC | 597 | Special Topics | 3.000 | 3.000 | A | 12.000 |
| Course Topic: | | POLITICAL GEOGRAPHY | | | | |
| PLSC | 597 | Special Topics | 3.000 | 3.000 | A | 12.000 |
| Course Topic: | | Machine Learning | | | | |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 4.000 | Term Totals | 10.500 | 10.500 | 9.000 | 36.000 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 4.000 | Comb Totals | 10.500 | 10.500 | 9.000 | 36.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 4.000 | Cum Totals | 19.500 | 19.500 | 18.000 | 72.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 4.000 | Comb Totals | 19.500 | 19.500 | 18.000 | 72.000 |

Summer 2021

Program: Liberal Arts
Plan: Political Science (PHD) Major

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| PLSC | 596 | Individual Studies | 5.000 | 5.000 | A | 20.000 |
| Course Topic: | | Gender in American Politics | | | | |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 4.000 | Term Totals | 5.000 | 5.000 | 5.000 | 20.000 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 4.000 | Comb Totals | 5.000 | 5.000 | 5.000 | 20.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 4.000 | Cum Totals | 24.500 | 24.500 | 23.000 | 92.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 4.000 | Comb Totals | 24.500 | 24.500 | 23.000 | 92.000 |

Robert A. Kubat

Robert A. Kubat, University Registrar

## Graduate Official Transcript

**Name:** **Chi Shun Fong** **SSN: ***-**-8313**

**Student ID:** **940461840** **Birthdate (MM/DD): 06/04**

### Fall 2021

Program: Liberal Arts
Plan: Political Science (PHD) Major
Plan: Social Data Analytics (NDUAL) Dual Title Program

Program: Liberal Arts
Plan: Political Science (MA) Major

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| PLSC | 504 | Topics in Pol Meth | 3.000 | 3.000 | A- | 11.010 |
| PLSC | 513 | Writing and Prof | 1.500 | 1.500 | P | 0.000 |
| PLSC | 550 | Com Pol Thry/Meth | 3.000 | 3.000 | A | 12.000 |
| PLSC | 597 | Special Topics | 3.000 | 3.000 | A | 12.000 |
| Course Topic: | | State Repression | | | | |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 3.890 | Term Totals | 10.500 | 10.500 | 9.000 | 35.010 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 3.890 | Comb Totals | 10.500 | 10.500 | 9.000 | 35.010 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 3.970 | Cum Totals | 35.000 | 35.000 | 32.000 | 127.010 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 3.970 | Comb Totals | 35.000 | 35.000 | 32.000 | 127.010 |

### Spring 2022

Program: Liberal Arts
Plan: Political Science (PHD) Major
Plan: Social Data Analytics (NDUAL) Dual Title Program

Program: Liberal Arts
Plan: Political Science (MA) Major

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| PLSC | 518 | Survey Design | 3.000 | 3.000 | B+ | 9.990 |
| PLSC | 555 | Comp Regimes | 3.000 | 3.000 | A | 12.000 |
| SODA | 501 | Big Social Data | 3.000 | 3.000 | A | 12.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 3.780 | Term Totals | 9.000 | 9.000 | 9.000 | 33.990 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 3.780 | Comb Totals | 9.000 | 9.000 | 9.000 | 33.990 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 3.930 | Cum Totals | 44.000 | 44.000 | 41.000 | 161.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 3.930 | Comb Totals | 44.000 | 44.000 | 41.000 | 161.000 |

### Summer 2022

Program: Liberal Arts
Plan: Political Science (PHD) Major
Plan: Social Data Analytics (NDUAL) Dual Title Program

Program: Liberal Arts
Plan: Political Science (MA) Major

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| PLSC | 596 | Individual Studies | 5.000 | 5.000 | A | 20.000 |
| Course Topic: | | Gender in American Politics | | | | |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 4.000 | Term Totals | 5.000 | 5.000 | 5.000 | 20.000 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 4.000 | Comb Totals | 5.000 | 5.000 | 5.000 | 20.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 3.930 | Cum Totals | 49.000 | 49.000 | 46.000 | 181.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 3.930 | Comb Totals | 49.000 | 49.000 | 46.000 | 181.000 |

### Fall 2022

Program: Liberal Arts
Plan: Political Science (PHD) Major
Plan: Social Data Analytics (NDUAL) Dual Title Program

Program: Liberal Arts
Plan: Political Science (MA) Major

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| GEOG | 586 | Geo Info Analysis | 3.000 | 3.000 | A | 12.000 |
| PLSC | 597 | Special Topics | 3.000 | 3.000 | A | 12.000 |
| Course Topic: | | Contentious Politics | | | | |
| SODA | 502 | Soc Data Analytics | 3.000 | 3.000 | A | 12.000 |
| STAT | 557 | Data Mining I | 3.000 | 3.000 | A | 12.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 4.000 | Term Totals | 12.000 | 12.000 | 12.000 | 48.000 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 4.000 | Comb Totals | 12.000 | 12.000 | 12.000 | 48.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 3.950 | Cum Totals | 61.000 | 61.000 | 58.000 | 229.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 3.950 | Comb Totals | 61.000 | 61.000 | 58.000 | 229.000 |

Robert A. Kubat

Robert A. Kubat, University Registrar

## Graduate Official Transcript

**Name:** **Chi Shun Fong**
**Student ID:** **940461840**

**SSN:** *****-**-8313**
**Birthdate (MM/DD):** **06/04**

### Spring 2024

Program: Liberal Arts
Plan: Political Science (PHD) Major
Plan: Social Data Analytics (NDUAL) Dual Title Program

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| PLSC | 600 | Thesis Research | 9.000 | 9.000 | R | 0.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 0.000 | Term Totals | 9.000 | 9.000 | 0.000 | 0.000 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 0.000 | Comb Totals | 9.000 | 9.000 | 0.000 | 0.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 3.950 | Cum Totals | 70.000 | 70.000 | 58.000 | 229.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 3.950 | Comb Totals | 70.000 | 70.000 | 58.000 | 229.000 |

### Summer 2024

Program: Liberal Arts
Plan: Political Science (PHD) Major
Plan: Social Data Analytics (NDUAL) Dual Title Program

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| PLSC | 596 | Individual Studies | 1.000 | 1.000 | R | 0.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 0.000 | Term Totals | 1.000 | 1.000 | 0.000 | 0.000 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 0.000 | Comb Totals | 1.000 | 1.000 | 0.000 | 0.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 3.950 | Cum Totals | 71.000 | 71.000 | 58.000 | 229.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 3.950 | Comb Totals | 71.000 | 71.000 | 58.000 | 229.000 |

### Fall 2024

Program: Liberal Arts
Plan: Political Science (PHD) Major
Plan: Social Data Analytics (NDUAL) Dual Title Program

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| PLSC | 601 | Ph D Dis Full-Time | 0.000 | 0.000 | R | 0.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 0.000 | Term Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 0.000 | Comb Totals | 0.000 | 0.000 | 0.000 | 0.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 3.950 | Cum Totals | 71.000 | 71.000 | 58.000 | 229.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 3.950 | Comb Totals | 71.000 | 71.000 | 58.000 | 229.000 |

### Spring 2025

Program: Liberal Arts
Plan: Political Science (PHD) Major
Plan: Social Data Analytics (NDUAL) Dual Title Program

| Course | | Description | Attempted | Earned | Grade | Points |
|---|---|---|---|---|---|---|
| PLSC | 601 | Ph D Dis Full-Time | 0.000 | 0.000 | R | 0.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Term GPA | 0.000 | Term Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Transfer Term GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined GPA | 0.000 | Comb Totals | 0.000 | 0.000 | 0.000 | 0.000 |

| | | | Attempted | Earned | GPA Units | Points |
|---|---|---|---|---|---|---|
| Cum GPA | 3.950 | Cum Totals | 71.000 | 71.000 | 58.000 | 229.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 3.950 | Comb Totals | 71.000 | 71.000 | 58.000 | 229.000 |

Graduate Career Totals

| | | | | | | |
|---|---|---|---|---|---|---|
| Cum GPA: | 3.950 | Cum Totals | 71.000 | 71.000 | 58.000 | 229.000 |
| Transfer Cum GPA | | Transfer Totals | 0.000 | 0.000 | 0.000 | 0.000 |
| Combined Cum GPA | 3.950 | Comb Totals | 71.000 | 71.000 | 58.000 | 229.000 |

### Non-Course Milestones

Doctoral Qualifying Examination
Status: Completed
Program: Liberal Arts
Date Completed: 11/17/2021

Culminating Experience: SCHOLARLY ESSAY
Status: Completed
Program: Liberal Arts
Date Completed: 11/01/2022

Doctoral Comprehensive Examination
Status: Completed
Program: Liberal Arts
Date Completed: 06/24/2024

End of Graduate Official Transcript

Robert A. Kubat
Robert A. Kubat, University Registrar

PennState

Office of the University Registrar

The Pennsylvania State University
112 Shields Building
University Park, PA 16802
(814) 865-6357
www.registrar.psu.edu

**RELEASE OF INFORMATION**
In compliance with the Family Education Rights and Privacy Act of 1974, this information is released on the condition that the recipient "will not permit any other party to have access to such information without the written consent of the student."

**OFFICIAL DOCUMENT**
"Penn State" appears in small print on a blue background across the entire face of the official academic transcript which is printed with security ink. The official transcript bears the seal of Penn State and the signature of the University Registrar. The raised University seal is not required.

**OFFICIAL ELECTRONIC DOCUMENT**
The official electronic transcript bears the seal of Penn State and the signature of the University Registrar. The blue ribbon security feature appears at the top of the transcript when opened in Adobe Reader.

**ACCREDITATION**
The Pennsylvania State University, whose prime purpose has always been to serve the people and the interests of the Commonwealth and the Nation, is accredited by the Middle States Association and is a member of the Association of American Universities.

Each of the Penn State Dickinson Schools of Law is on the approved list of the American Bar Association and is a member of the Association of American Law Schools.

**INSTITUTIONAL CODE**
The institutional code for The Pennsylvania State University is 003329.

**TRANSCRIPTS KEYS**
To view academic transcript keys for all academic careers, please visit our website at http://www.registrar.psu.edu/transcripts/transcript_key_index.cfm

PennState

Department of Political Science
College of the Liberal Arts
The Pennsylvania State University
401 Susan Welch Liberal Arts Building
137 Fischer Road
University Park, PA 16802

814-865-7515
Fax: 814-863-8979
polisci.la.psu.edu

1 September 2026

To Whom It May Concern:

I write in my capacity as Director of Graduate Studies in the Department of Political Science at Penn State University to certify that Chi Shun “Gary” Fong has satisfied all requirements of our doctoral program, including the oral defense of his dissertation, and will be awarded his diploma in December 2026, when Penn State next holds graduation ceremonies.

If you require any additional information, please do not hesitate to contact me.

Cordially,

Douglas Lemke

Douglas Lemke
Professor and Director of Graduate Studies

E-mail: dwl14@psu.edu

# Theories, Methods, and Practices in Computational Social Science

**Course Overview**

Social science research is now regularly conducted using data that are larger and more complex than the data for which conventional statistical tools were designed. Examples of such data include individual-level consumption records, data streams collected from social media, and archives of electronic/digitized government records. These fine-grained, voluminous, and high-dimensional data require a set of methods that are more sophisticated than the conventional toolkits of quantitative social science, which would allow us to learn patterns from the data, select variables with predictive power, reduce the dimensionality of data by grouping variables, and make predictions based on these variables. In this course, we will learn the basic principles of machine learning methods and appreciate how they are applied in solving social science problems. This course focuses on textual data – one of the most common data types in mainstream computational social science research.

**Course Objectives**

1. Develop a basic understanding of the theory behind machine learning methods
2. Understand how to explore, describe, and visualize data
3. Apply machine learning methods to real-life data, with a particular focus on text
4. Appreciate how machine learning methods are applied to solve social science problems

**Prerequisites**

This course is an intensive course for advanced undergraduate and postgraduate students interested in machine learning and automated text analysis. Students enrolled in this course are expected to have basic knowledge of statistics, quantitative methods, and programming. Students who would like to develop their future research agenda on computational social science are especially welcome.

**Assessment**

- 10% - In-class participation
    - Active participation is required
- 40% - Individual coding exercise x 4
    - Solving specific tasks with methods learned in class
- 50% - Group Project (5% research prospectus; 15% presentation; 30% final report)
    - 2 – 3 students form a group
    - A group project that applies the methods learned in this class to solve an issue in students' own disciplines

**Course Structure**

The weekly class will be divided into three sessions of 45 minutes each. The first session covers the basic theoretical concepts and walks through the application of relevant methods in the assigned readings (if any). The second and third sessions are two labs in which all students follow the instructor to run programming code and do in-class coding exercises.

**Weekly Schedule and Readings**
The schedule below gives the issues covered by the first session, the content of the lab sessions, and the required readings and applications each week. The readings and the applications (focus on how different methods are applied) listed should be read before class that day. The full citations can be found below in the references section. Students are expected to spend, on average, 3 hours per week reading the assigned materials and working on coding.

<u>Part A Introduction to Machine Learning, R and Text Analysis</u>

**Download R and Install R Studio before the start of the semester*

*Week 1 Introduction*
- Issues: intuition and background of machine learning, prediction vs. explanation, the online community of Data Science
- Lab: basic setup of RStudio, the basic operation of the R language in RStudio, data type and structure, condition and iteration
- Reading: Rhys (2020) Ch1; Shmueli et al. (2010)

**Apply for Youtube API, Install chromium*

*Week 2 Data wrangling and visualization*
- Issues: split-apply-combine/map-reduce, common charts and plots in social science
- Lab: using packages, read and write files, data wrangling with the *dplyr*, data visualization with *ggplot*, RMarkdown
- Reading: Rhys (2020) Ch2; Wickham & Grolemund (2016) Ch3 & Ch27

*Week 3 Problems with Big Data and data collection*
- Issues: problems with Big Data, how API works, web-scrapping
- Lab: retrieve YouTube data using API, web-scrape data with R*Selenium*
- Reading: Salganik (2017) Ch1 & Ch2
- Application: Teo & Fu (2021)

**HW 1 due*

*Week 4 Text as Data and Text (pre)-processing*
- Issues: text as data, bag-of-word as text representation, text-preprocessing, sparsity & hashing, Zipf's Law
- Lab: regular expression, quanteda package: preprocessing, document-feature matrix (dfm), tokens' co-occurrence
- Reading: Grimmer & Brandon (2013); Grimmer et al. (2022) Ch5
- Application: Luqiu & Lu (2021),

*Week 5 Text visualization, dictionary-based analysis, and basics of classification*
- Issues: bias-variance trade-off, labeling data, inter-coder reliability (ICR), cross-validation, performance metric

- Lab: word cloud and network graph, sentiment analysis, coding a training set, ICR Test
- Reading: Rhys (2020) Ch3; Silge & Robinson (2017) Ch8
- Application: Rheault et al. (2019)

Part B Supervised Methods: Classification and Regression

*Week 6 Naïve Bayesian and Logistics Regression*
- Issues: parametric methods: pros and cons, (naïve) Bayesian rule, Logistic Regression
- Lab: Naïve Bayesian, Logistics Regression
- Reading: Rhys (2020) Ch4 & Ch6 (6.1-6.3)
- Application: Chenoweth & Ulfelder (2017)

**HW 2 due*

*Week 7 Support Vector Machine (SVM), Decision Tree, and Random Forest,*
- Issues: non-parametric methods: pros and cons, hyper-tuning, SVM, Decision Tree, Ensemble Methods like Random Forest
- Lab: SVM, Random Forest, XGboost
- Reading: Rhys (2020) Ch6 (6.4 and after), Ch7 & Ch8
- Application: Pan (2019)

**Research prospectus due*

*Week 8 Regression and Regularization*
- Issues: basic of regression, regularization: lasso, ridge, and elastic net
- Lab: Regression in machine learning, regularization
- Reading: Rhys (2020) Ch9 & Ch11
- Application: Mitts (2019)

Part C Unsupervised Methods: Dimension Reduction

*Week 9 Principal Component Analysis (PCA) and Clustering*
- Issues: PCA, various approaches to clustering
- Lab: PCA, clustering
- Reading: Rhys (2020) Ch4 13, Ch16 &17
- Application: Li (2022)

**HW 3 due*

*Week 10 Topic Modeling*
- Issues: Generative topic modeling, topic modeling with structure
- Lab: Latent Dirichlet Allocation (LDA), Structural Topic Modeling (STM)
- Reading: Silge & Robinson (2017) Ch6
- Application: Urman et al. (2021)

Part D Advanced Topic

*Week 11 Neural Network*

- Issue: Deep learning, basics of the neural network, Recurrent Neural Network (rNN), Convolutional Neural Network (cNN)
- Lab: Neural Network with non-textual data
- Reading: Günther & Fritsch (2010); Lectures 1 from "Introduction to Deep Learning." MIT: http://introtodeeplearning.com/
- Application: Zhang & Pan (2019)

*HW 4 due

*Week 12 Natural Language Process (NLP), Word-embedding, Transfer learning*

- Issue: NLP, word-embedding as text representation, transfer-learner: Bert
- Lab: using Bert in R
- Reading: Rodriguez & Spirling (2022); Jon Harmon – RBert - https://www.youtube.com/watch?v=1kV0fPDefBM&ab_channel=LanderAnalytics
- Application: Dai (2018)

*Week 13 Generative AI*

- Issue: Current development of Generative AI, prompt engineering, AI ethics
- Lab: Local use of large language model with ollama
- Reading: Palmer & Spirling (2023)
- Application: Rettenberger et al. (2025)

*Week 14 Presentation*

***Reference***

Chenoweth, E., & J. Ulfelder (2017). Can structural conditions explain the onset of nonviolent uprisings? *Journal of Conflict Resolution 61*(2), 298-324.

Dai, Y. (2018). Measuring populism in context: A supervised approach with word embedding models. *Working Paper*.

Grimmer, J and Brandon S. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Documents. *Political Analysis, 21*(3): 267-297.

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Günther, F. and S. Fritsch (2010). neuralnet: Training of neural networks. *The R journal, 2*(1), 30-38.

Li, L. (2022). Decoding political trust in China: A machine learning analysis. *The China Quarterly, 249*, 1-20.

Luqiu, L. R., & Lu, S. (2021). Bounded or boundless: A case study of foreign correspondents' use of twitter during the 2019 Hong Kong protests. *Social Media+ Society*, *7*(1), 2056305121990637.

Mitts, T. (2019). From isolation to radicalization: anti-muslim hostility and support for ISIS in the west. *American Political Science Review 113* (1), 173-194.

Pan, J. (2019). How Chinese officials use the internet to construct their public image. *Political Science Research and Methods,* 7 (2), 197-213.

Palmer, A., & Spirling, A. (2023). Large language models can argue in convincing ways about politics, but humans dislike AI authors: implications for governance. *Political science*, *75*(3), 281-291.
Rettenberger, L., Reischl, M., & Schutera, M. (2025). Assessing political bias in large language models. *Journal of Computational Social Science*, *8*(2), 1-17.
Rheault, L., Rayment, E., & Musulan, A. (2019). Politicians in the line of fire: Incivility and the treatment of women on social media. *Research & Politics*, *6*(1), 2053168018816228.
Rhys, H. (2020). *Machine Learning with R, the tidyverse, and mlr*. Simon and Schuster
Rodriguez, P. L., & Spirling, A. (2022). Word embeddings: What works, what doesn't, and how to tell the difference for applied research. *The Journal of Politics, 84*(1), 101-115.
Salganik, Matthew J. (2017). *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.
Shmueli, G. (2010). To explain or to predict?. *Statistical science, 25*(3), 289-310.
Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. O'Reilly.
Teo, E., & Fu, K. W. (2021). A novel systematic approach of constructing protests repertoires from social media: comparing the roles of organizational and non-organizational actors in social movement. *Journal of Computational Social Science*, *4*(2), 787-812.
Urman, A., Ho, J. C. T., & Katz, S. (2021). Analyzing protest mobilization on telegram: The case of 2019 anti-extradition bill movement in Hong Kong. *Plos one*, *16*(10), e0256675.
Wickham, H., & Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. O'Reill.
Zhang, H., & Pan, J. (2019). CASM: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociological Methodology, 49*(1), 1-57.

# Politics and Development of Contemporary China

**Course Overview**

The experience of the People's Republic of China (PRC) has invited scholarly attention because of its stable one-party rule accompanied by rapid economic growth that began more than three decades ago. To what extent is the China experience unique? To what extent can the China experience be explained and analyzed by existing theories in political science? In what ways can the China experience contribute to the scholarship of comparative politics? This course focuses on the comparative implications of the China case. While its substantive focus is on China, the course situates the case of China within the broad literature of comparative politics and international relations, including state-building, the politics of development, contentious politics, and authoritarian politics. It is organized along the following four sections:

1. *A Political History of Modern China*
   Why did China become a communist state in the 20th century? How can we understand political changes in contemporary China (1949-present)? This section reviews China's political history, including late imperial China, the nationalist revolution, and the communist revolution, focusing on issues like state-building, nation-building, and the leadership of the Chinese Communist Party.

2. *Political Institutions and Development*
   How can we explain China's state structure and the development of its domestic political institutions? What are the political-economic explanations for China's economic miracle? This section introduces China's domestic political institutions, including formal state institutions, central-local relations, state-business interactions, and external trade.

3. *Public Opinion and Political Participation*
   How active and influential are societal actors in China's political landscape? What are Chinese citizens' views on politics? What are the main social cleavages in China? This section explores key topics in China's political sociology, including state-society relations, the rise and decline of NGOs, and major social cleavages such as gender, labor, environmental protection, and minority concerns.

4. *China and the World*
   As a historical global power and (re)emerging global power in the contemporary era, China has developed complex relations with its neighbors and other countries. This section examines China's foreign policy, regional security dynamics, its role in global governance, and its relationships with neighboring states, the Global South, and the Global North.

**Course Objective**

1. Understand and explain political outcomes in China with social science theory
2. Situate China within the broader context of comparative politics literature and debates.

**Assessments**

- 10% - Attendance and participation

    - Active participation is required
    - Each student is required to lead the discussion for one seminar
- 20% x2 – Response papers
    - Write a 1000-word response paper to the readings of one research cluster (Part A, B, C, and D)
    - Can submit at any time during the term
- 50% – Research Proposal (5% research prospectus; 15% presentation; 30% final report)
    - A research proposal with a research question, a literature review, hypotheses, and methods

**Suggested Data Sources for China Studies**

- China Gazetteer (This database aggregates over 3000 county gazetteers compiled since the founding of the People's Republic of China).
- CCP Elites - China Data Lab - UC San Diego (This data contains information on the highest members of China's government, including all full and alternate Central Committee members, provincial standing committee members, and vice governors from 1997 through July of 2021.)
- National Population Census of China Database (This database provides access to digitized reports and Excel files of the population census published by China Statistics Press.)
- Enterprise Survey for Innovation and Entrepreneurship in China (Targeting private and foreign-owned enterprises registered in China from 2010 to 2017, this survey includes a total of 58,500 corporations and individual household businesses in 117 counties and districts).
- Asian Barometer Survey (The ABS is a cross-national comparative survey focused on socioeconomic modernization, regime transition and democratization, and changes in political values across the East Asia region.)
- China General Society Survey (CGSS) 2023 (Launched in 2003, the CGSS is the country's earliest nationwide, comprehensive, and ongoing academic survey project. It systematically and comprehensively collects data at multiple levels, including social, community, household, and individual levels.)

**Weekly Schedule and Readings**

Part A Historical Legacy

*Week 1: The birth of Modern China: Late Qing to the Early Communist Era*

Hui, V. T. B. (2025). What Rose and Fell in Imperial China? State Strength, the Civil Service Examination and Inventiveness. *Political Science Quarterly*, qqaf013.

Fiskesjö, M. (2006). Rescuing the empire: Chinese nation-building in the twentieth century. *European Journal of East Asian Studies*, *5*(1), 15-44.

Fitzgerald, J. (1990). The Misconceived Revolution: State and Society in China's Nationalist Revolution, 1923–26. *The Journal of Asian Studies*, *49*(2), 323-343.

Strauss, J. (2006). Morality, coercion and state building by campaign in the early PRC: Regime consolidation and after, 1949–1956. *The China Quarterly*, *188*, 891-912.

**Watch documentary: China: Century of Revolution (Part 1)*

*Week 2: The Communist Party and Chinese Leadership*

Lieberthal, Kenneth. (2003). *Governing China: from revolution through reform*. New York: W. W. Norton. (Chapter 7)

Manion, M. (1985). The cadre management system, post-Mao: The appointment, promotion, transfer and removal of party and state leaders. *The China Quarterly*, *102*, 203-233.

Meyer, D., Shih, V. C., & Lee, J. (2016). Factions of different stripes: gauging the recruitment logics of factions in the reform period. *Journal of East Asian Studies*, *16*(1), 43-60.

Thornton, P. M. (2021). Party all the time: The CCP in comparative and historical perspective. *The China Quarterly*, *248*(S1), 1-15.

<u>Part B Political Institutions and Development</u>

*Week 3: Chinese State Institutions*

Landry, P. F. (2008). *Decentralized Authoritarianism in China: The Communist Party's control of local elites in the post-Mao era*. Cambridge University Press. (Chapter 2)

O'brien, K. J., & Han, R. (2009). Path to democracy? Assessing village elections in China. *Journal of Contemporary China*, *18*(60), 359-378.

Potter, P. B. (1999). The Chinese legal system: Continuing commitment to the primacy of state power. *The China Quarterly, 159*, 673-683.

Truex, R. (2014). The returns to office in a "rubber stamp" parliament. *American Political Science Review*, *108*(2), 235-251.

*Week 4: Chinese Political Economy I: Explaining Growth*

Bosworth, B., & Collins, S. M. (2008). Accounting for growth: comparing China and India. *Journal of economic perspectives*, *22*(1), 45-66.

Heilmann, S. (2008). Policy experimentation in China's economic rise. *Studies in comparative international development*, *43*(1), 1-26.

Shevchenko, A. (2004). Bringing the party back in: the CCP and the trajectory of market transition in China. *Communist and Post-Communist Studies*, *37*(2), 161-185.

Xu, C. (2011). The fundamental institutions of China's reforms and development. *Journal of Economic Literature*, *49*(4), 1076-1151.

*Week 5: Chinese Political Economy II: State-Business Relations*

Jiang, J. (2018). Making bureaucracy work: Patronage networks, performance incentives, and economic development in China. *American Journal of Political Science*, *62*(4), 982-999.

Lin, S. (2025). Addressing Risk by Doing Good: Business Response to Government Policy Initiative. *The Journal of Politics*, *87*(4), 1414 – 1430.

Lorentzen, P., Landry, P., & Yasuda, J. (2014). Undermining authoritarian innovation: the power of China's industrial giants. *The Journal of Politics*, *76*(1), 182-194.

Ong, L. H. (2012). Between developmental and clientelist states: Local state-business relationships in China. *Comparative Politics*, *44*(2), 191-209.

*Week 6: Chinese Political Economy III: Politics of International Trade*

Erten, B., & Leight, J. (2021). Exporting out of agriculture: The impact of WTO accession on structural transformation in China. *Review of Economics and Statistics*, *103*(2), 364-380.
Fuchs, A. (2018). China's economic diplomacy and the politics-trade nexus. In *Research Handbook on Economic Diplomacy* (pp. 297-314). Edward Elgar Publishing.
Gu, J., Zhang, C., Vaz, A., & Mukwereza, L. (2016). Chinese state capitalism? Rethinking the role of the state and business in Chinese development cooperation in Africa. *World Development*, *81*, 24-34.
McNally, C. A. (2012). Sino-capitalism: China's reemergence and the international political economy. *World politics*, *64*(4), 741-776.

<u>*Part C Public Opinion and Political Participation*</u>

*Week 7: Authoritarian Control: Repression, Cooption and Legitimation*
King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American political science Review*, *107*(2), 326-343.
Wang, Y., & Minzner, C. (2015). The rise of the Chinese security state. *The China Quarterly*, *222*, 339-359.
Xu, X. (2021). To repress or to co-opt? Authoritarian control in the age of digital surveillance. *American Journal of Political Science*, *65*(2), 309-325.
Zhu, Y., & Fu, K. W. (2024). How propaganda works in the digital era: Soft news as a gateway. *Digital Journalism*, *12*(6), 753-772.

**Research prospectus due*

*Week 8: State-Society Relations I: Political Participation and Political Responsiveness*
Fu, D., & Distelhorst, G. (2018). Grassroots participation and repression under Hu Jintao and Xi Jinping. *The China Journal, 79*(1), 100-122.
O'brien, K. J. (1996). Rightful resistance. *World Politics*, *49*(1), 31-55.
Qiaoan, R., & Teets, J. C. (2020). Responsive authoritarianism in China - a review of responsiveness in xi and Hu administrations. *Journal of Chinese Political Science*, *25*(1), 139-153.
Su, Z., & Meng, T. (2016). Selective responsiveness: Online public demands and government responsiveness in authoritarian China. *Social science research*, *59*, 52-67.

*Week 9: State-Society Relations II: NGOs and Civil Society*
Dai, J., & Spires, A. J. (2018). Advocacy in an authoritarian state: How grassroots environmental NGOs influence local governments in China. *The China Journal*, *79*(1), 62-83.
Fu, D. (2017). Disguised collective action in China. *Comparative Political Studies, 50*(4), 499-527.
Hsu, J. Y., & Hasmath, R. (2014). The local corporatist state and NGO relations in China. *Journal of Contemporary China*, *23*(87), 516-534.
Saich, T. (2000). Negotiating the state: The development of social organizations in China. *The China Quarterly*, *161*, 124-141.

*Week 10: State-Society Relations III: Major Social Cleavages*

Liu, C. (2019). Local public goods expenditure and ethnic conflict: Evidence from China. *Security Studies*, *28*(4), 739-772.
White, Tyrene. 2010. "Domination, Resistance and Accommodation in China's One-Child Campaign." In *Chinese Society*, 3rd ed. Routledge.
Wu, A. X., & Dong, Y. (2019). What is made-in-China feminism (s)? Gender discontent and class friction in post-socialist China. *Critical Asian Studies*, *51*(4), 471-492.
Steinhardt, H. C., & Wu, F. (2016). In the name of the public: Environmental protest and the changing landscape of popular contention in China. *The China Journal*, *75*(1), 61-82.

*Part D China and the World*

*Week 11: China and the World: Overview*
Callahan, W. A. (2008). Chinese Visions of World Order: Post-Hegemonic or a New Hegemony? *International Studies Review*, *10*(4), 749-761.
Haug, S., Foot, R., & Baumann, M. O. (2024). Power shifts in international organisations: China at the United Nations. *Global Policy, 15*, 5-17.
Mattingly, D., Incerti, T., Ju, C., Moreshead, C., Tanaka, S., & Yamagishi, H. (2025). Chinese state media persuades a global audience that the "China model" is superior: Evidence from a 19-country experiment. *American Journal of Political Science*, *69*(3), 1029-1046.
Shambaugh, D. (2015). China's soft-power push: The search for respect. *Foreign Affairs*, *94*(4), 99-107.

*Week 12: China and the Global North*
Christie, Ø. S., Jakobsen, J., & Jakobsen, T. G. (2024). The US way or Huawei? An analysis of the positioning of secondary States in the US-China rivalry. *Journal of Chinese Political Science, 29*(1), 77-108.
Liao, S., & McDowell, D. (2016). No reservations: International order and demand for the renminbi as a reserve currency. *International Studies Quarterly*, *60*(2), 272-293.
Medeiros, E. S. (2019). The changing fundamentals of US-China relations. *The Washington Quarterly*, *42*(3), 93-119.
Reilly, J. (2017). China's economic statecraft in Europe. *Asia Europe Journal*, *15*(2), 173-185.

*Week 13: China and the Global South*
Callahan, W. A. (2016). China's "Asia Dream" the belt road initiative and the new regional order. *Asian Journal of Comparative Politics*, *1*(3), 226-243.
Dreher, A., Fuchs, A., Parks, B., Strange, A. M., & Tierney, M. J. (2018). Apples and dragon fruits: The determinants of aid and other forms of state financing from China to Africa. *International Studies Quarterly*, *62*(1), 182-194.
Eichenauer, V. Z., Fuchs, A., & Brückner, L. (2021). The effects of trade, aid, and investment on China's image in Latin America. *Journal of Comparative Economics*, *49*(2), 483-498.
Hooijmaaijers, B. (2021). China, the BRICS, and the limitations of reshaping global economic governance. *The Pacific Review*, *34*(1), 29-55.

*Week 14 Presentation*

**Borrowing Credibility to Shape Public Opinion:**
**Western Influencers as Chinese Government Foreign Defenders**

Chi Shun (Gary) Fong
PhD Candidate, Department of Political Science, Pennsylvania State University

**Abstract**

Recent advances in digital communication enable authoritarian regimes to shape the perception of democratic audiences and soften their demands for punitive foreign policies against human rights violations. While state mouthpieces like Russia Today and CGTN actively push state narratives online, their rigid propaganda style and explicit foreign attribution limit both their persuasive appeal and their likelihood of being selected by skeptical audiences. I argue that seemingly independent foreign social media influencers—acting as credible ingroup storytellers—allow autocrats to bypass these institutional barriers. I evaluate this argument through a preregistered survey experiment exposing U.S. respondents (N = 1,044) to pro-Chinese government YouTube videos addressing human rights in Xinjiang, manipulating both **primary source origin** (state media vs. influencer creation) and **intermediary endorser** (state media vs. influencer sharing). Results show that while pro-Chinese government narrative content from both sources reduces respondents' objections to Chinese state actions in Xinjiang, state media endorsement degrades perceived source credibility, reducing overall persuasiveness and discouraging information selection. By identifying influencers as gateways for foreign propaganda across both persuasion and selection stages, this study uncovers a key mechanism of cross-border opinion manipulation.

**Keywords:** foreign propaganda, influencers, authoritarianism, social media

## Introduction[1][2][3]

While the advancement of the internet and communication technology has challenged authoritarian regimes by increasing information flows and boosting collective action potential (Zhuravskaya et al., 2020), recent studies suggest authoritarian regimes with strong digital capacity are well equipped to handle these challenges (Keremoğlu & Weidmann, 2020; Tucker et al., 2017) or even deploy these technologies to manipulate public opinion both domestically (Zhu & Fu, 2023) and abroad (Brady, 2015). Nonetheless, extant research on digital repression explores how authoritarian regimes manipulate domestic public opinion using propaganda (D. C. Mattingly & Yao, 2022), censorship (Y. Chen & Yang, 2019; King et al., 2013), and distraction and diversion (Roberts, 2018). But this research pays little attention to authoritarian information operations abroad.

Like domestic propaganda, foreign propaganda targeting audiences abroad shapes autocratic survival (Dukalskis, 2021). On the one hand, foreign propaganda helps authoritarian regimes promote their foreign policies and helps foster an international environment conducive to authoritarianism. For instance, Chinese state media and diplomats engage more frequently and positively on social media with countries that have stronger economic ties to China (Liang, 2021; Villegas-Cruz, 2024b). Similarly, the Russian government disseminates misinformation in the West through coordinated efforts by state media and bots (Bennett & Livingston, 2018) to disrupt democratic politics (Eady et al., 2023; Elswah & Howard, 2020) and exacerbate polarization (Bradshaw et al., 2023). Further, foreign propaganda neutralizes negative narratives from foreign and exiled critics who may incite domestic opposition (Escribà-Folch & Wright, 2015; Nugent & Siegel, 2024) and pressures democratic states to intervene (Tomz et al., 2020; Tomz & Weeks, 2020). Notable examples include the Chinese government's information strategies to counter international opprobrium from the outbreak of Covid-19 by, for example, highlighting shared victimhood (Loh & Loke, 2024) and medical aid from China (Villegas-Cruz, 2024a). One of the CCP's largest foreign information operations aims to defend itself against

[1] This study has passed the ethical reviews of Penn State (STUDY00025171)

[2] This study is supported by the Penn State's Department of Political Science, Program in Comparative Politics (PiCP) research fund

[3] This study's hypotheses, design, and analysis plan were preregistered on the Open Science Framework (OSF)

alleged human rights violations in Xinjiang by featuring religious extremism and its benevolent rule in Xinjiang (Alpermann & Malzer, 2024; Villegas-Cruz, 2025).

Two gaps in the literature persist, however. First, extant research mostly focuses on state actors – such as state media, diplomats, and governmental agencies – who are typically unfamiliar to foreign audiences. As a result, we know little about the predominant way foreign audiences consume propaganda, namely through non-state actors, such as social media influencers (Brockling et al., 2023). Social media influencers are broadly defined as third-party actors who gain influence over niche groups through social media activities (Hudders et al., 2021). Though these influencers are non-state actors, they are increasingly coordinating information strategies with governments in both democracies and autocracies (Wang & Zhang, 2025; Woolley, 2022). Therefore, I examine how Western social media influencers, as non-state actors, deploy foreign propaganda and how their information shapes attitudes among foreign audiences. Second, while some research finds that foreign propaganda prompts foreign citizens to adopt narratives and policy positions aligned with the interests of authoritarian regimes (Carter & Carter, 2021; Kobayashi et al., 2025; Mattingly et al., 2024), others show that foreign information campaigns are often not credible and suffer from low engagement with foreign audiences (Crilley et al., 2022; Henriksen et al., 2024; Liang et al., 2023; Min & Luqiu, 2021). These studies lack consensus, in part, because we do not yet understand the key mechanisms that make foreign propaganda effective.

I argue that the persuasiveness of foreign propaganda depends on the perceived credibility of the information source. In the modern online information environment, credibility is multi-layered. Unlike in the age of printed and broadcasting media, online information is processed through various stages before reaching the audience. As such, the credibility of online information depends on factors beyond the primary source's origin, such as intermediary endorsers, i.e., who is sharing the information with you (Mena et al., 2020; Metzger et al., 2010; Shen et al., 2019). In this context, propaganda messages produced or shared by social media influencers, who are not only better storytellers than state sources but also convey a higher level of credibility, could better persuade foreign citizens. This is because influencers appear neutral and convey a sense of authenticity, which makes the content seem less propagandistic. More importantly, influencers share critical ingroup identities with target audiences, such as nationality, race, and cultural background. Social identity theory proposes that people treat those they perceive as ingroup members more favorably than outgroup members, which, in turn, influences individuals' attitudes and behavior (Huddy, 2001). This high perceived credibility, whether it originates from the source or the endorser, not only mediates the persuasive effect of the message but also could extend the reach of propaganda (Brockling et al., 2023) by guiding selective information consumption (Messing & Westwood, 2014; Shehata & Strömbäck, 2022).

I test this theory using a preregistered survey experiment exposing U.S. respondents to pro-Chinese government videos publicly available on YouTube. The experimental design manipulates the video's primary source origin (a Western influencer vs. China Global Television Network [CGTN], a Chinese state media outlet) alongside additional intermediary endorsers (a Western influencer's endorsement of CGTN's video and CGTN's endorsement of a Western influencer's video) to isolate how distinct credibility signals from the two sources affect audience attitudes toward Chinese state actions in Xinjiang and support for U.S. punitive actions targeting the Chinese government. Results show that while pro-Chinese government video content generally reduces opposition to Chinese state actions in Xinjiang, explicit state media endorsement triggers a significant credibility penalty that diminishes persuasive impact and depresses the audience's information selection. This influencer advantage was especially pronounced among respondents skeptical of the mainstream media and with prior negative views of China. However, exposure and endorsement of either source did not shift respondents' support for U.S. punitive actions targeting the Chinese government.

This paper advances the literature on foreign information strategies by theorizing how information sources influence information credibility, with implications for understanding the persuasiveness of foreign information in high-choice media contexts. In these information environments, overt state branding undermines message persuasiveness, whereas seemingly organic influencer cues enhance it. This research demonstrates that influencers could be effective information brokers not only because they could persuade foreign audiences but also because they could better reach targeted audiences. These dynamics highlight authoritarian resilience, in which information strategies soften perceptions among ordinary democratic citizens rather than elites, which may reduce their support for domestic and diasporic anti-regime activism, as well as punitive foreign policy, in the long run. Further, because audiences cannot easily distinguish influencer-mediated propaganda from organic discourse, governments face acute regulatory challenges, forcing democratic platforms to balance labeling requirements with free expression, as autocracies exploit fragmented attention economies to broaden their external support networks.

**Effects of Foreign Propaganda Messages: Mixed Evidence**

Few studies explore how autocratic propaganda targets foreign citizens. Some find that foreign propaganda is most persuasive on salient issues about which audiences have little prior knowledge and pre-existing opinions. For instance, Carter & Carter (2021) demonstrate that exposure to Russian-sponsored propaganda shifts U.S. citizens' opinions on less salient issues, such as the U.S. role in global affairs, rather than on more salient domestic issues. Similarly, Mattingly et al. (2024) find that exposure to propaganda from the external media arms of the Chinese and U.S. governments moves citizens' preferences toward their respective governance models, which is also an abstract issue. Other studies perceive foreign propaganda as a form of misinformation. Mader et al. (2022) find that German voters who mistrust the government, believe in conspiracy theories, and are less politically informed are vulnerable to Russian narratives. Yet Kobayashi et al. (2025) suggest the opposite: that Japanese respondents are generally susceptible to Chinese official narratives.

In addition to issue salience and audience attributes, the perceived credibility of information messages shapes their persuasiveness. Credibility, which entails reasonable grounds for being believed (Meyer, 1988), has long been established as central to making messages persuasive (Pornpitakpan, 2004). Propaganda content that blatantly adopts a pro-regime position and is inconsistent with mainstream narratives (Mader et al., 2022) is unlikely to be perceived as credible or persuasive in a society with a free media environment. For instance, Min & Luqiu (2021) find that standard propaganda justifying one-party rule in China was perceived as less credible to foreign audiences than a propaganda video praising Chinese meritocracy. While state-sponsored media outlets can learn to produce content that better fits foreign contexts (Brady, 2015), using official government sources remains a problem for enhancing messages' perceived credibility (Druckman, 2001; Ladd, 2010; Miller & Krosnick, 2000). Compared with information from domestic mainstream media, propaganda content from unfamiliar external sources is perceived as less credible. Dai and Luqiu (2020), for example, find that the same Chinese propaganda content was more persuasive to U.S. citizens when it appeared in well-known Western media outlets than in Chinese outlets. Similarly, Mader et al. (2022) find that Russian propaganda is only effective among German citizens who distrust the mainstream media.

Credibility further limits the potential effects of foreign propaganda by guiding media consumption. The classic information reception model (e.g., Zaller, 1992 and Geddes & Zaller, 1989) reminds us that the effect of a propaganda message is contingent on exposure. In communication research, credibility, in both online and offline settings, plays a dual role: conditioning the persuasive effects of information and guiding the selection of information sources (Messing & Westwood, 2014; Shehata & Strömbäck, 2022). Yet citizens living in democracies, unlike their counterparts in authoritarian regimes, enjoy a wide range of media choices and are not compelled to be exposed to authoritarian propaganda. Although advances in internet and communication technologies increase the global reach of foreign propaganda, it remains unclear whether foreign citizens select information from pro-authoritarian sources rather than from competing or opposing narratives. For instance, Henriksen et al. (2024) find that content from RT and Sputnik, two Russian state-backed media outlets, has minimal reach even within European domestic alternative news ecosystems. Crilley et al. (2022) discovered that Twitter followers of RT rarely engaged with RT's content and also followed other mainstream news channels. The impact of perceived credibility on the circulation of foreign propaganda is more clearly demonstrated by Liang et al. (2023), who found that sharing tweets from China's state media accounts, particularly political tweets, dropped immediately and continued to decline after these accounts were labeled as state-affiliated.

**Borrowing Credibility from Influencers for Political Propaganda**

I argue that authoritarian regimes leverage the credibility of social media influencers to amplify their foreign propaganda. Social media influencers can be broadly defined as third-party actors who gain opinion-shaping power over a niche group of people through social media activities

(Hudders et al., 2021). By showcasing their expertise (Wasike, 2023), narrating with attractive and fashionable linguistic styles (Huffaker, 2010; Schmuck et al., 2022), and developing virtual social relationships with followers (Cheng et al., 2024; Harff, 2022), they become credible information sources for their audiences. Further, objective and visual indicators, such as verification badges and the number of likes, shares, mentions, and subscriptions (Treem et al., 2020), increase the perception of their influence. Recent survey evidence illustrates the scope of their online influence: the most active 25% of U.S. adults account for more than 95% of posts on Twitter (McClain et al., 2021) and TikTok (Bestvater, 2024). Given their influence and increasing popularity on social media, they now resemble traditional opinion leaders who serve as information brokers and gatekeepers, mediating mass communication (Lazarsfeld & Katz, 1955; Zaller, 1992). In that sense, they provide credibility to information not only as primary sources but also as intermediary endorsers. This endorsing role is even more critical in online settings, as internet users seldom directly access information from primary production sources. Instead, online users' exposure to information heavily depends on other factors, such as platform algorithms and social reaccommodation from intermediaries (Messing & Westwood, 2014). For instance, Metzger et al. (2010) found that endorsements from peers can outweigh initial skepticism about an unfamiliar information source.

Influencers address everyday life issues, such as fashion, sports, or gaming, and leverage their sway for commercial purposes, while also becoming increasingly politicized in both democracies and authoritarian regimes (Riedl et al., 2023). In democracies, some social media influencers engage in daily online political discussions (Wang & Zhang, 2025) and electoral campaigns (Shmargad, 2022), posting content that resembles partisan narratives. While some influencers may genuinely hold their expressed partisan positions, many are driven by material and non-material incentives under the auspices of marketing firms and political parties (Goodwin et al., 2023; Woolley, 2022). Compared with traditional party-led political advertisements, influencers convey authenticity by producing personalized, ostensibly nonpartisan content that is perceived as more credible and appealing than official state media and direct communication from politicians' social media feeds. Other influencers remain independent of governments, shaping followers' political attitudes and behaviors, including political interests (Harff, 2022), online and offline political participation (Cheng et al., 2024), and political efficacy (Wasike, 2023). In authoritarian regimes, despite censorship, credible public opinion leaders — such as critical journalists, intellectuals, and activists — can challenge official propaganda (Nip & Fu, 2016) and facilitate protest mobilization (Poell et al., 2016), much like their counterparts in democracies. Yet authoritarian regimes with strong digital capacity, such as the CCP, both recognize this threat and exert a firmer grip on the behavior of online opinion leaders (Creemers, 2017). For instance, Chinese celebrities are compelled to repost content from Chinese government agencies' social media accounts to survive in the entertainment industry, which is tightly controlled by the government (D. Chen & Gao, 2023).

The advantages of incorporating social media influencers into domestic political advertising and propaganda campaigns apply to foreign authoritarian governments as well. Recent studies show that both the Russian and Chinese governments hire social media influencers to influence public opinion in democracies (Brockling et al., 2023; Woolley, 2022), which helps build a positive image of these authoritarian regimes in the minds of foreign audiences. Moreover, I posit that this effect translates into policy influence, shaping how democratic societies perceive and respond to authoritarian regimes. Citizens' moral evaluation of authoritarian regimes, alongside other factors such as threat perception, shapes not only citizens' support for punitive policies against such regimes (Arı & Sonmez, 2025; M. R. Tomz & Weeks, 2020) but also the positions of policymakers who observe their constituents' preferences (M. Tomz et al., 2020). Therefore, propaganda messages produced or endorsed by foreign influencers, compared to those from official state media institutions, will be more effective in persuading foreign citizens to adopt favorable attitudes toward an authoritarian regime and lower their support for punitive foreign policies **(H1).**

**H1 (Persuasive Effect of Influencer Propaganda)**: Foreign citizens are more effectively persuaded by propaganda produced or endorsed by foreign influencers than that by official institutions of authoritarian regimes to:

**a)** adopt a more favorable attitude toward the authoritarian regime.

**b)** lower support for punitive actions against the authoritarian regime

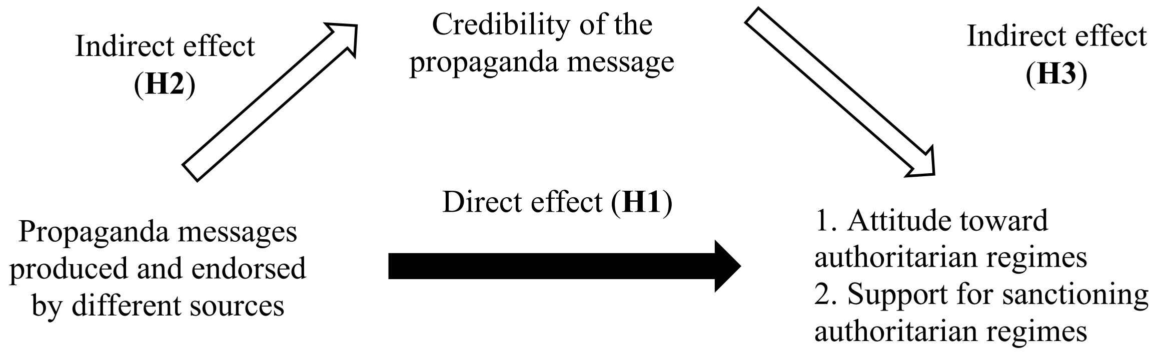
Content produced by social media influencers, compared to that produced by the official institutions of authoritarian regimes, better aligns with foreign contexts and internet culture, as influencers are natives of those online spaces and may adopt a more personal communication style and a softer presentation. These qualities make their messages more authentic. Thus, influencer-produced content is more convincing than both traditional didactic party-state content and the authoritarian propagandists' broadcasting style, which persists in their foreign propaganda (Fan et al., 2024; Min & Luqiu, 2021). This is especially true when foreign influencers reside in and frequently travel throughout authoritarian regimes, enabling them to produce content that appears apolitical and close to the lives of ordinary people.

More importantly, foreign citizens are likely to perceive foreign influencers as more credible than official government communication. Influencers not only sell the autocrats' story by sharing authentic and personal experiences but do so while concealing their affiliation with these institutions. Further, a common ingroup identity, shared by both the influencer and the audience, strengthens the apparently neutral image. Audio and visual cues indicating shared ingroup identity include race and ethnicity as well as accent and tone. A consistent finding in social identity theory suggests that people treat those they perceive as ingroup members more favorably, which, in turn, influences political attitudes and behavior (Huddy, 2001). Similarly, in public diplomacy, ingroup messages are perceived as more credible and persuasive by foreign audiences because insider communication is likely to be perceived as more agreeable and constructive than messages from outgroup members (Abramson et al., 2025; Dragojlovic, 2015). Building on this logic, I expect that foreign citizens will perceive a message produced or endorsed by a foreign influencer as more credible than one from official institutions of authoritarian regimes (**H2**), and that this perception will, in turn, positively mediate the persuasive effect of propaganda (**H3**).

**H2 (Credibility of Influencer Propaganda)**: Foreign citizens perceive propaganda messages produced or endorsed by foreign influencers as more credible than those from official institutions of authoritarian regimes.

**H3 (Mediation Effect of Propaganda Credibility)**: The perceived credibility of the propaganda message positively mediates the persuasive effect of the propaganda message in **H1**.

**Fig. 1 Causal diagram**




Finally, foreign influencers enjoy a higher level of perceived credibility, an antecedent variable guiding information selection in both offline (Shehata & Strömbäck, 2022) and online settings (Messing & Westwood, 2014). Thus, I posit that a foreign influencer is more likely than an official institution of authoritarian governments to be selected by foreign citizens when they seek additional information about China (H4).

**H4 (Selection Effect of Influencer Endorsement)**: Foreign citizens are more likely to select a foreign influencer than an official institution of authoritarian governments as an information source.

**Research Design**

I recruited a nationally representative sample of U.S. citizens in April 2025 for a preregistered survey experiment that randomly assigned respondents to view edited, real pro-Chinese government videos available on YouTube from the Chinese state media outlet China Global Television Network (CGTN) or Western pro-Chinese government influencers. These videos adopt narratives that defend the Chinese government on the controversial issue of human rights in Xinjiang. Among the treatment groups, some respondents received an additional social media profile of the two sources as endorsers of the videos. By manipulating these cues of credibility of the primary production source and intermediary endorser, I tested whether perceived credibility mediates the persuasive effect of pro-Chinese government videos on respondents 'attitudes towards the Chinese government's actions in Xinjiang and their support for dovish and hawkish punitive actions toward the Chinese government. In a pretreatment question, I also asked whether the respondents prefer a social media profile showing a Western influencer living in China or CGTN as their information source about China.

*Case selection*

I study the case of Xinjiang for several reasons. First, preliminary evidence indicates that the Chinese government works with Western influencers on social media to promote the Chinese version of the Xinjiang story – "borrowing mouths from foreign friends," as Chinese officials responsible for foreign propaganda have described it (Ryan et al., 2021, 2022). Second, human rights issues strongly motivate democratic citizens to support their governments in adopting more proactive policies against authoritarian regimes (M. R. Tomz & Weeks, 2020), thereby giving authoritarian states clear incentives to sway foreign public opinion on this issue. Thus, this case is a prototypical example of authoritarian foreign propaganda. Third, human rights in Xinjiang receive extensive coverage in Western democracies, including the U.S.[4] As discussed above, propaganda is less likely to persuade citizens on issues when they are more informed. If influencer-assisted Chinese propaganda proves effective in this case, it could demonstrate its capacity to influence these democratic publics. Finally, the U.S. government is increasingly hostile toward China and has become a leading force behind international sanctions aimed at China. Understanding how Chinese propaganda influences U.S. public attitudes and policy positions has important implications for U.S. foreign policy and China's long-term development.

*Treatments selection*

In treatment groups, respondents either watched a YouTube video anchored by a Chinese journalist from CGTN or by a Western white influencer. I selected real videos to increase the experiment's external validity. Randomly assigning respondents to receive a video from the pool reduces bias by minimizing the possibility that the treatment videos are unrepresentative or that respondents are pre-treated (Carter & Carter, 2021). I selected videos from CGTN because it is the largest foreign-facing propaganda apparatus of the Chinese government, providing content in

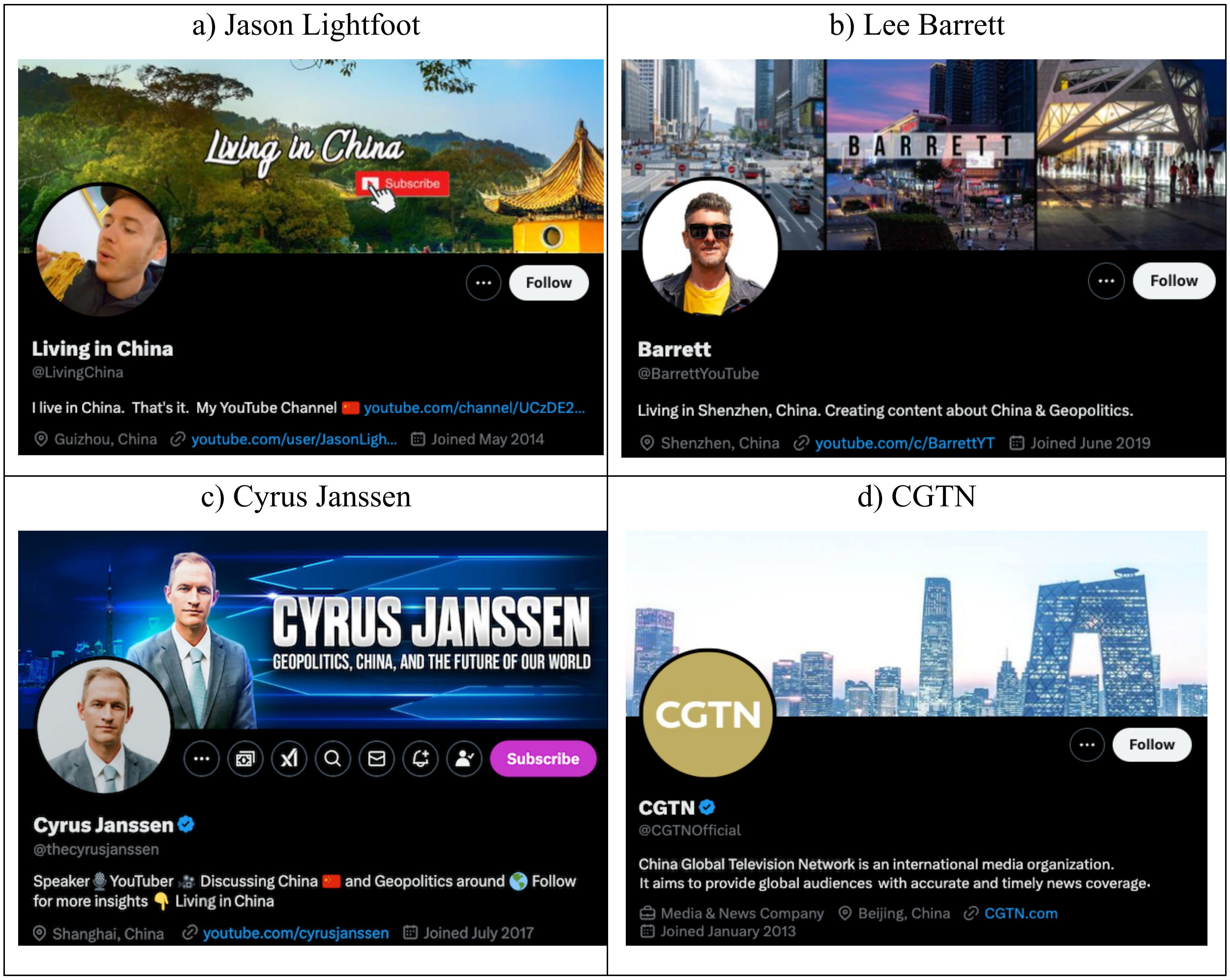
[4] For instance, a poll by YouGov in September 2022 says that 57% of the adult U.S citizens have heard about the treatment of the Uyghur people by the Chinese government, and 53% of them believe the treatment qualifies as a human rights violation – see https://today.yougov.com/international/articles/43714-americans-think-china-guilty-human-rights-abuses

all six UN official languages. Following Mattingly et al. (2024), I selected videos by analyzing the text description of CGTN's YouTube videos about Xinjiang.[5] For influencers' videos, I curated a list of pro-Chinese government influencers based on the earlier studies of Brockling et al. (2023) and other policy research briefs (Ryan et al., 2021, 2022; Seitz et al., 2022) on this issue, following a list of criteria that ensure their White identity, pro-Chinese government positions, and online influence, and selected videos that match the narratives of CGTN's video.[6] There are three pairs of treatment videos from both sources on three common narratives used by the Chinese government on the Xinjiang issue (Alpermann & Malzer, 2024) – (1) praising the government's efforts in developing Xinjiang, (2) the prevalence of terrorism in Xinjiang under external influences, and (3) cultural and religious freedom enjoyed by the ethnic minorities in Xinjiang. This helps control differences in content between CGTN's and influencers' videos. Additionally, all videos were edited to around 1.2 to 1.5 minutes and included subtitles to minimize potential biases from differences in exposure duration and the speaker's English accent.[7] With these measures in place, I minimized differences in the main message of the video and highlighted differences in the identities of the video anchors and the sources.

Some respondents also received the actual profile of either CGTN or one of the three influencers on X after watching the video and being told that this profile shared the video to create an impression of endorsement. There are three common heuristics from non-content attributes that social media users use to evaluate the credibility of a message's source (Lin et al., 2016), including authority, bandwagon cues, and identity. I standardized these profiles according to the first two heuristics. Regarding authority, all profiles indicated that the account holders are informed about China, either as foreigners living in China or a Chinese media outlet. For bandwagon cues, I removed all subscription counts to prevent respondents from evaluating the profiles based on their perceived influence. The major difference among these profiles was their identity. While the CGTN profile revealed its connection to China through its full name, the influencers' profiles featured a white person's face and bios typical of influencers.

[5] See Appendix A1 for the analysis
[6] See Appendix A2 for the criteria
[7] See Appendix D for the treatment videos' transcripts and snapshots.

**Fig. 2. Profile of White Influencers and CGTN on X**

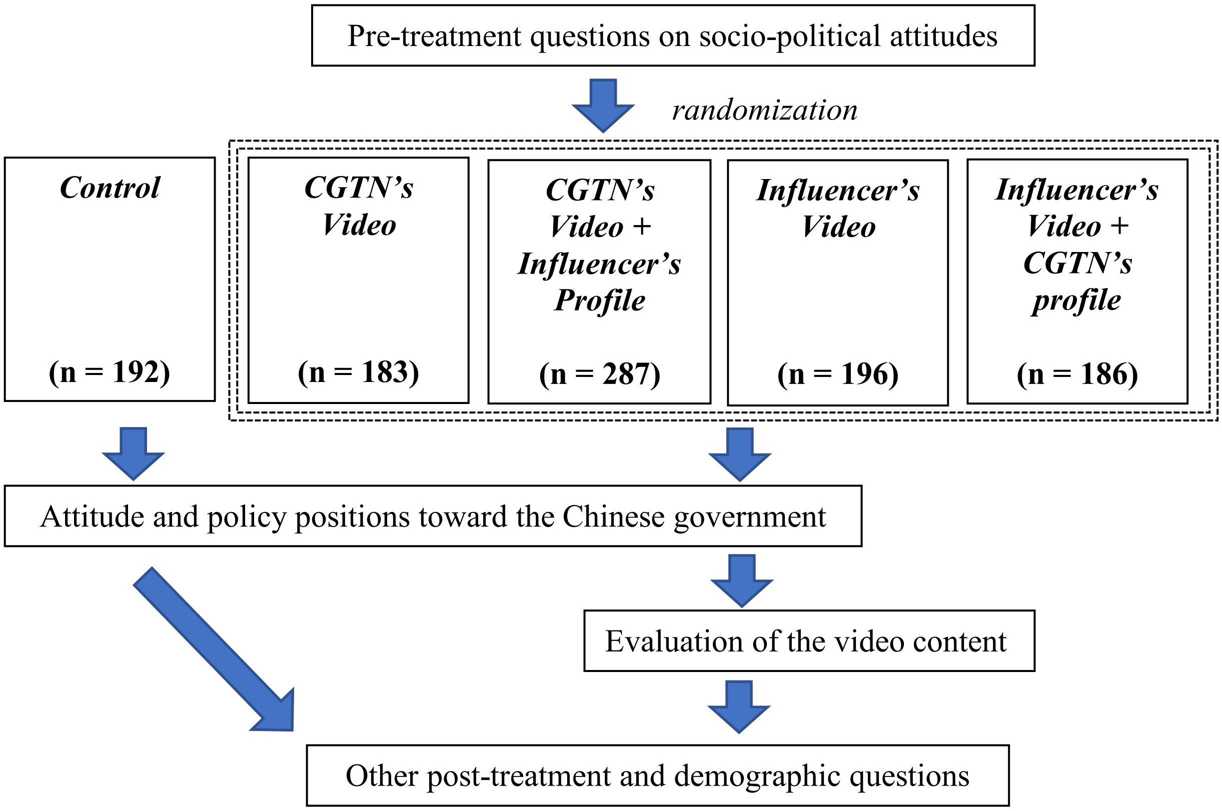



*Experimental Design*

1,098 U.S. citizens were recruited as respondents via Cint Theorem, and the results of 1,044 who passed an instructional manipulation check were included in the study. Missing data are assumed to be missing at random and imputed using multiple imputation with the *MICE* package in *R*. All respondents first answered a set of pre-treatment questions on socio-political attitudes and behavior, including being asked to choose either CGTN's social media profiles or one of the influencers as their preferred source of information about China.[8] As these profiles were also used in some treatment groups, this question was administered well before the treatment to avoid contamination of its results and those of the outcome questions.

[8] See Appendix E for the survey questions.

**Fig. 3. Flow of survey**




Then, respondents were randomly assigned to one of the five experimental conditions: (1) a placebo video showing natural scenery (*Control*), (2) one video from CGTN (*CGTN's Video)*, (3) one CGTN's videos with one influencers' profiles (*CGTN's Video + Influencers' Profile*), (4) one videos from influencers (*Influencers' Video)*, and (5) one influencers' video with CGTN's profile (*Influencers' Video + CGTN's profile*).[9] While *CGTN's Video* and *Influencers' Video* serve to test the persuasive effects of these videos, the additional *CGTN's Profile* and *Influencers' Profile* examine whether their endorsements moderate the effects of the treatment videos from the other source.

I did not treat respondents with the video and profile from the same source for two reasons. First, having pure video treatments helped to isolate the effect of the videos. Second, there were enough cues in the videos to suggest their sources, including the ethnicity of the

[9] Alternatively, this study could be perceived as having 6 video treatment groups or having 18 sub-treatment groups if we consider the status of the profile treatment. Thus, the treatment condition *CGTN's Video + Influencers' Profile* had a higher sample size to allow each of the 9 sub-treatment groups to have a comparable size with the others.

narrators in the videos, style of the narration, and a short sentence on each survey page of the treatment videos – *Please watch this video produced by [a China-based media organization the China Global Television Network (CGTN)]*, or [*a social media influencer living in China, the name of the influencer]*. Thus, profiles from the same source provide little additional information to respondents, and the treatment effect associated with an endorser is already partially captured by the videos. Furthermore, the profile was displayed after the video with a pause. This design isolated the profile's endorser effect from the video content by allowing respondents, as in other treatment groups, to first watch the videos and then evaluate them after the pause more independently of the content based on the new information provided.

After the treatment, all respondents were asked to answer three sets of primary outcome questions that are increasingly aggressive towards China, including (1) objections to the Chinese government's action in Xinjiang*,* (2) support of the U.S government taking dovish actions against the Chinese government, and (3) support of the U.S government taking hawkish actions against the Chinese government. I used (1) as a proxy for respondents' perceptions of the Chinese government, as democratic citizens tend to associate authoritarian regimes with human rights violations (Barceló & Sheen, 2024). Moreover, these treatment videos are explicitly created to defend the Chinese government on the Xinjiang issue. Thus, directly asking respondents for their views on the Xinjiang issue better captures the video's effects. For (2) and (3), I followed the literature and selected four actions as proxies for the dovish (public condemnation and economic sanctions) and hawkish actions threatening and using military force) (Mattingly & Sundquist, 2023; Xu, 2025). I averaged the two actions for each policy type. Only respondents from the treatment groups were asked to assess the videos' credibility in presenting the situation in Xinjiang and the Uyghurs. These three outcome questions were measured using a 5-point Likert scale. After answering these outcome questions, all respondents completed additional post-treatment and demographic questions.

*Identification*

The primary estimand is the average treatment effect (ATE) of watching a video endorsed and produced by Western white Influencers or CGTN on the primary outcomes (1) – (3) for **H1a** and **H1b**. For the primary outcomes, the pre-registered regression model is as follows:

$$Y_{ij} = \beta_0 + \beta_1 \, video + \beta_2 \, profile + \varphi X_i + \varepsilon_i$$

In the formula, (*i*) indexes the respondents, (*j*) indexes the three sets of primary outcome questions, (is the result from the primary outcome questions, *video* and *profile* are factored variables indicating (*i's*) treatment status of receiving which video or profile, ( $\beta_1$)is the effect of the video, ( $\beta_2$) is the effect of the social media profile additive to the video, $\varphi X_i$ and is a vector of binarized covariates. They include demographic characteristics such as *age, gender, race, education,* and political attitudes, including *partisanship, ideology, and political interest.* Additionally, I include two measures of respondents' orientation toward China, which are likely to affect the two primary outcomes and mediate the effect of the treatment: their *impressions of the Chinese government* and their *knowledge about Chinese politics and history*. I did not directly ask respondents about their knowledge of Xinjiang issues to avoid priming or post-treatment bias; instead, I asked two indirect but related pretreatment questions (Blackwell et al., 2025). Finally, $\varepsilon_i$ is the error term.

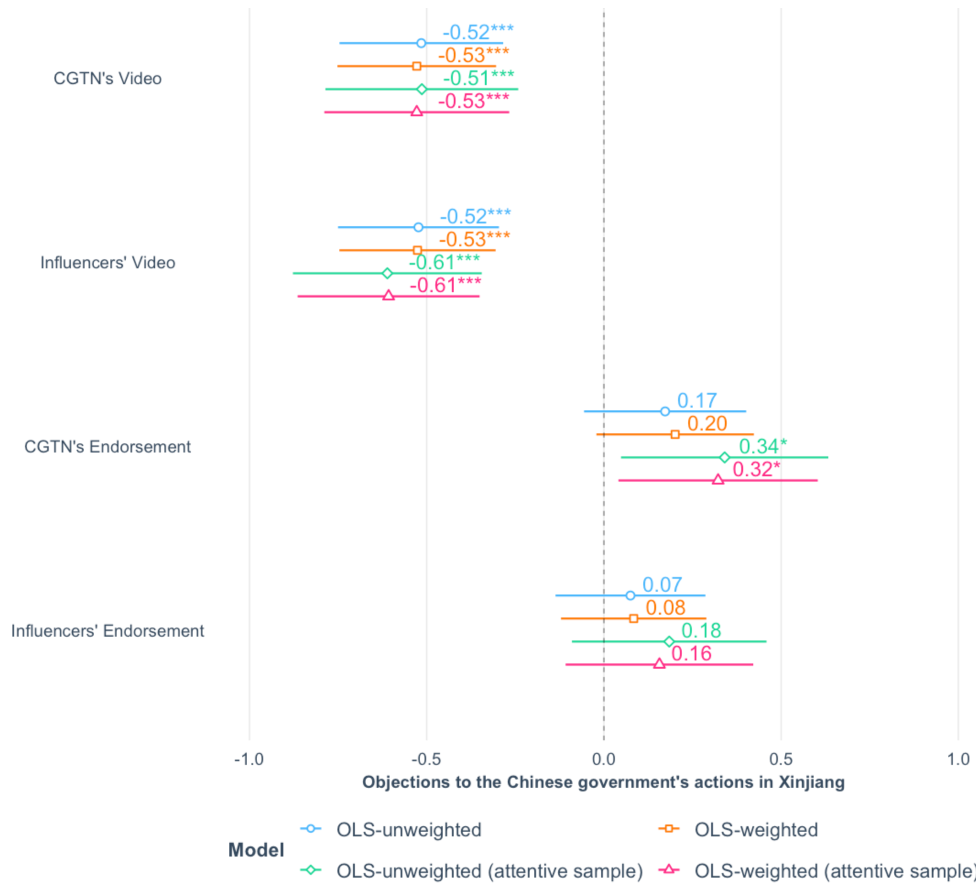
I employed four ordinary least squares (OLS) models for three primary outcomes to test **H1**, using the full sample and the sample that passed a treatment-related factual manipulation check, with or without covariate adjustment[10]. I tested **H2** by rerunning the same four models with the secondary outcome, but I changed the baseline comparison group to respondents from CGTN's video, since respondents from the control group were not asked to evaluate the video's content. To test **H3**, I used the model from Imai et al. (2011) to separate the indirect effect of credibility from the direct effect on respondents' objection to the Chinese government's actions in Xinjiang. In the appendix, I proved that my results are robust to alternative specifications by treating individual videos and profiles as the independent variables.[11] Finally, to test H4, I used a simple binomial test to see whether respondents pick CGTN's or a Western influencer's social media profile as their preferred source for additional information about China.

**Persuasive Effects of Pro-Chinese Government Content**

[10] Factual manipulation checks (FMCs) ask respondents objective questions about the experiment and better confirm attentiveness to the experiment than the instruction manipulation check. See Kane and Barabas (2019)
[11] See Appendix C

Fig. 4 presents the average treatment effects (ATE) on objections to the Chinese government's actions in Xinjiang.[12] The x-axis shows different treatment conditions, including the two primary sources of pro-Chinese-government videos, the additional intermediary endorsement of CGTN on the influencers' videos, and influencers' endorsements of CGTN's video. A larger value on the y-axis indicates a more critical attitude toward the Chinese government.

**Fig.4 ATE on objections to the Chinese government's actions in Xinjiang**




Notes: * p < 0.05, ** p < 0.01, *** p < 0.001.

In Fig. 4, both CGTN's and influencers' videos produced a strong and statistically significant negative effects (-0.5 to -0.6) on respondents' objections to China's actions in Xinjiang across models. In the two covariate-adjusted models, this magnitude is close to that of the positive effect of having a positive prior-impression of the Chinese government in the

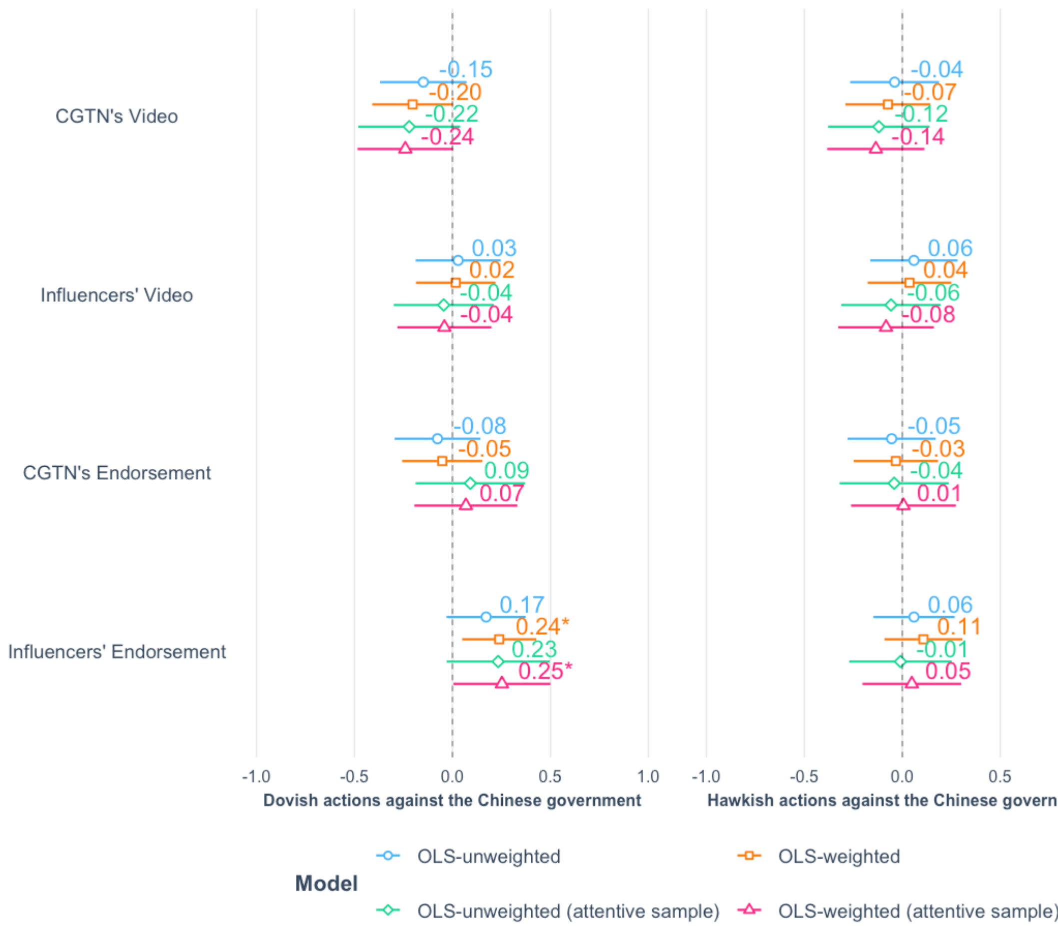
[12] See Appendix B3 for the tabulated results for Fig. 4 to Fig. 6.

full regression model (0.5). In other words, the results suggest that even brief exposure to pro-China content could offset the negative impression of China typically held by an average U.S. citizen. This finding challenges earlier claims that China's foreign propaganda is ineffective at shifting public opinion (Brady, 2015), given the Chinese government's persistent efforts to modernize its foreign propaganda, especially by cloaking it in a neutral tone (Moore & Colley, 2024). Among the attentive respondents, influencers' videos enjoyed a slightly larger persuasive effect (-0.6) than CGTN's video (-0.5), though the difference is not statistically significant.

Nonetheless, the figure also highlights the effect of endorsement cues via social media profiles. Endorsement of CGTN alongside an influencer's video produced a modest positive effect (0.3) among attentive samples, offsetting about 53% to 56% of the videos' persuasive effects. These results suggest that when U.S. respondents detect a clear Chinese affiliation in pro-China content on social media, the message's persuasiveness drops sharply. In contrast, exposure to an influencer's profile alongside a CGTN video had no effect, likely because respondents perceived these influencers as ordinary users when they were neither followers of these influencers nor able to assess the perceived influence of these users after the relevant cues were removed from the profiles. In this regard, influencers, even when perceived as ordinary social media users, enjoy a comparative advantage in disseminating pro-Chinese government content by avoiding the discredit of being associated with the Chinese government, thereby explaining their success in extending the reach of propaganda produced by Chinese state media (Brockling et al., 2023). Together, these findings provide partial evidence supporting **H1a.**

Fig. 5 shows the ATE on support for the U.S. government taking dovish and hawkish punitive actions against the Chinese government. The results show that videos and additional endorsement cues from both sources have almost no effect, except for small positive effects on support for dovish policies caused by influencer endorsements in the covariate-adjusted model. This suggests a backfiring effect, as respondents may believe that dovish actions, as the only viable foreign policy options currently on the table, should not be shaped by information from random social media users. Taken together, these results reveal that respondents' prior beliefs about appropriate diplomatic responses to the Chinese government remain stable, especially on costly military measures against a seemingly strong authoritarian regime, such that a once-off exposure, though changed their attitude to the issue *per se*, would not directly their policy position. This is consistent with the literature, which argues that factors beyond moral evaluations—such as perceived threat—also critically shape foreign policy preferences (M. Tomz et al., 2020; M. R. Tomz & Weeks, 2020). Overall, the results reject **H1b.**

**Fig.5 ATE on support for punitive actions against the Chinese government**

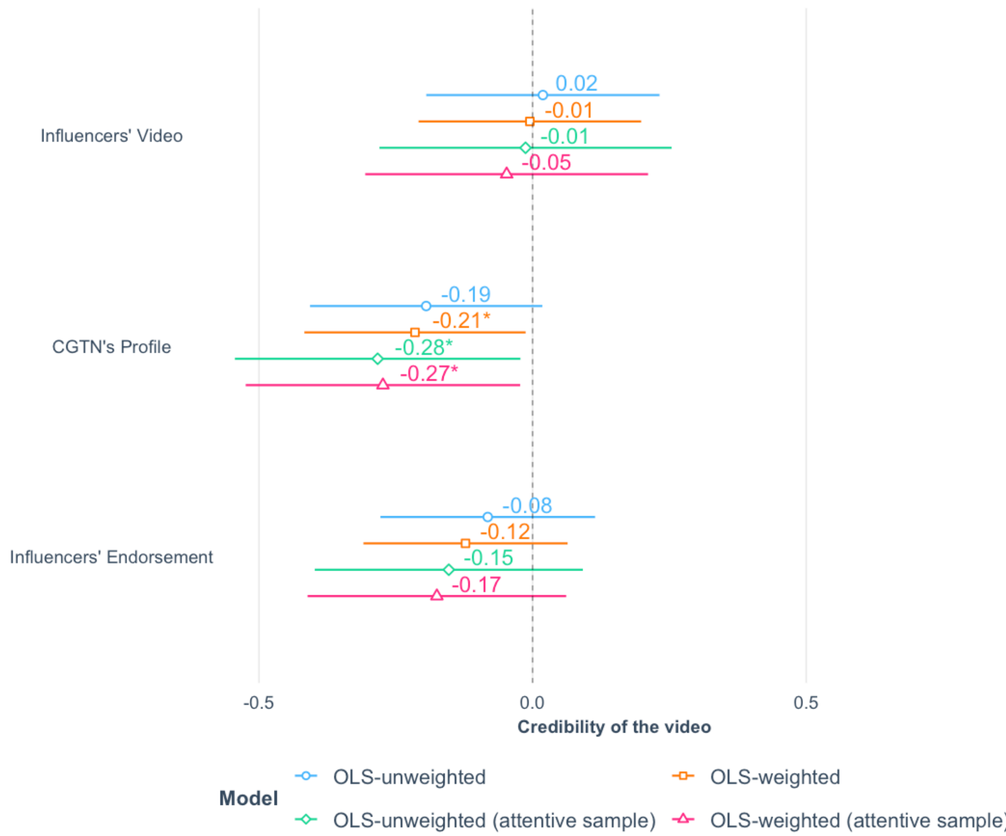



Notes: * p < 0.05, ** p < 0.01, *** p < 0.001.

## How Perceived Credibility Mediates the Persuasiveness of Pro-Chinese Government Content

Next, I examine respondents' perceptions of the treatments' credibility. In Fig. 6, respondents who received CGTN's video treatment served as the reference group because those in the control group were not asked to evaluate the video.

**Fig. 6. Credibility of pro-Chinese government videos from different sources and with different endorsement cues**

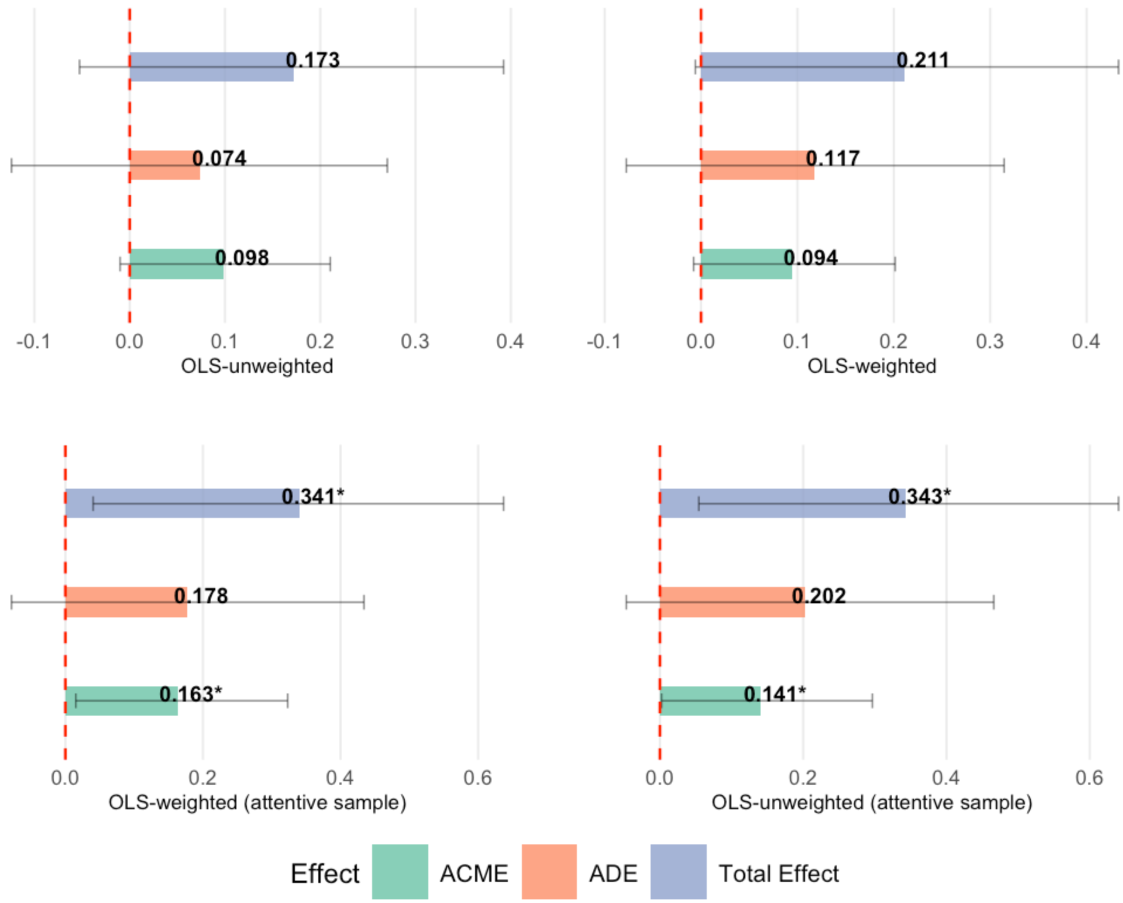



Note: * p < 0.05, ** p < 0.01, and *** p < 0.001.

Across models, influencers' videos showed no significant credibility advantage over CGTN videos, consistent with their persuasive effects on objections to the Chinese government's actions in Xinjiang as presented in Fig. 4. It is also unsurprising that the endorsement of influencers had no statistically significant effect on the credibility of the pro-Chinese government videos, again, given that respondents were neither followers of these influencers nor able to assess the online influence of these influencers. In contrast, CGTN's endorsement, which clearly signaled that it is a China-based media outlet, had modest negative effects (-0.2 to -0.3) on the video's credibility in most models. The results again provide partial support for **H2**, as CGTN's endorsement undermines the credibility of pro-Chinese government videos.

To test **H3,** which posits that the perceived credibility of the treatment mediates its persuasive effects, I employed a mediation analysis within the framework of Imai et al. (2011). This approach provides causal estimates of the extent to which a single treatment condition affects the outcome through the mediator using Monte Carlo draws for nonparametric bootstrap simulations. In mediation analysis, the ATE is decomposed into two components: the Average Causal Mediation Effect (ACME), the indirect treatment effect through the mediator, and the Average Direct Effect (ADE), the direct effect of the treatment. Here, the treatment that I focused on is the additional endorsement of CGTN for respondents who received the influencers' videos, as this is the only treatment condition that consistently shows statistically significant effects on the respondents' objections to the Chinese government's action in Xinjiang, our primary outcome, and the mediator of perceived credibility of the treatment videos.

**Fig. 7. Persuasive effects of CGTN's endorsement mediated by credibility**




Note: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Fig. 7 presents the mediation analysis results from 5,000 bootstrap simulations.[13] Across four models, the indirect effects of CGTN's endorsement were modest but comparable to the direct effect. More importantly, these effects were at least marginally significant, with p-values ranging from 0.03 to 0.075, whereas all the direct effects were insignificant in all cases after mediation, unlike the results in Fig. 4. These patterns indicate a classic partial mediation in which credibility is the mechanism behind the persuasive effects of CGTN's endorsement. Substantively, CGTN's endorsement reduced the persuasive effects of the videos in reducing respondents' objection to the Chinese government actions in Xinjiang presented in Fig. 4 by 18 to 27% through reduced credibility, providing partial evidence to support **H3**.

[13] See Appendix B4 for the tabulated regression results. The results are robust to unmeasured confounding between the mediator and the outcome variable, as shown by sensitivity tests in Appendix C2.

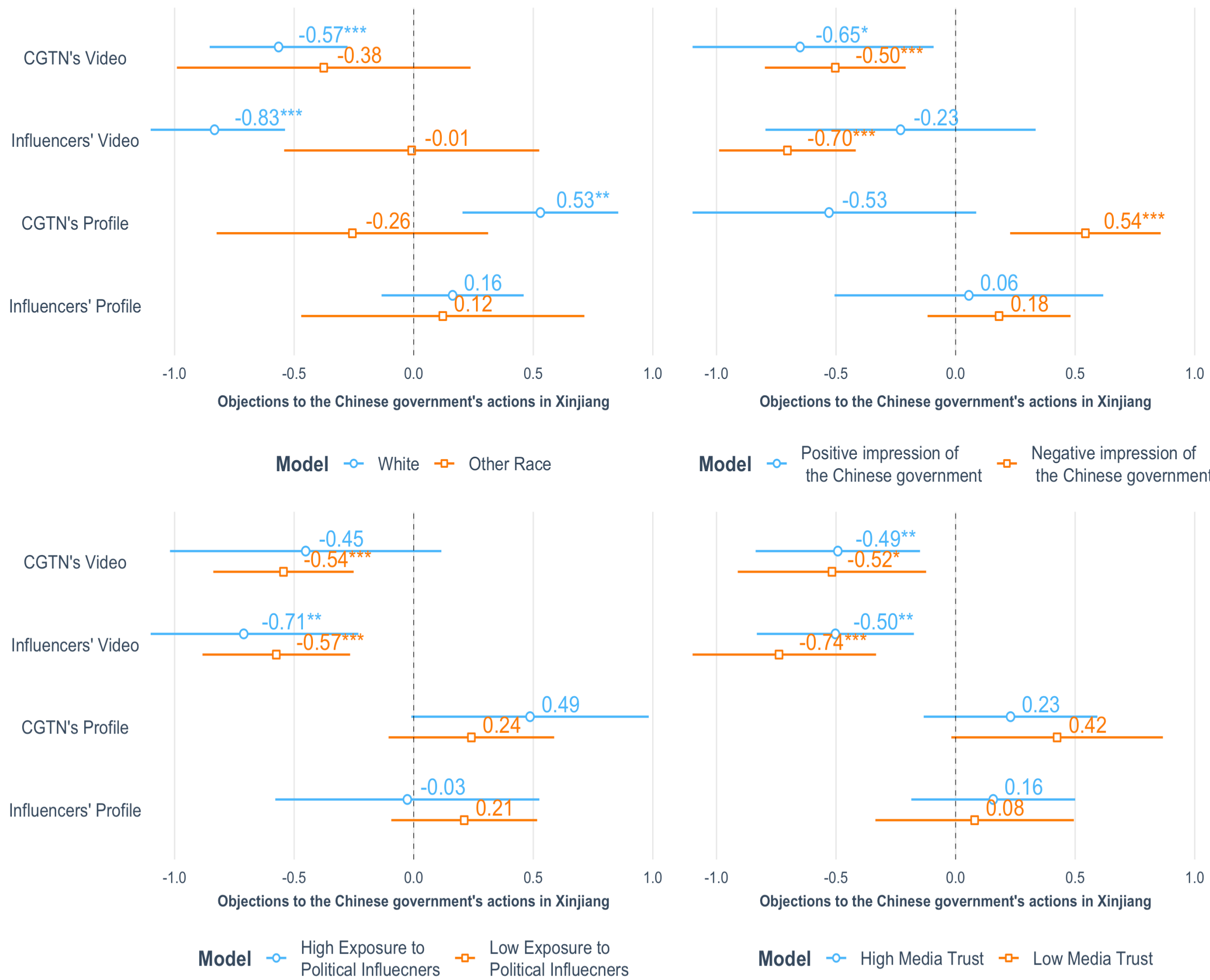
A point worth discussing is that, in both the full and attentive samples, the indirect effect did not fully capture the total effect of CGTN’s endorsement. There are two possible reasons. First, respondents’ perceived credibility of the videos may be overestimated for various reasons, such as acquiescence bias when answering questions about a less informed topic (Krosnick, 1991). Thus, the gap in perceived credibility between respondents who receive only the influencers’ videos and those who also receive the CGTN’s profile is reduced. Second, the presence of CGTN’s profile may have prompted respondents to evaluate the pro-Chinese government videos on qualities distinct from credibility and not captured by the credibility question, such as expectations, perceived authenticity, and interest in the videos. For instance, cueing respondents that the treatment video is from a Chinese media outlet may make it seem less relevant to them, thereby encouraging satisficing (Krosnick, 1991). Nonetheless, the insignificant direct effect in Fig. 7 suggests that an alternative mechanism is less likely to account for the remaining unexplained direct effect of CGTN’s endorsement.

**Targeted Persuasion and Selection Effects of Influencer-Endorsed/Created Content**

In this section, I explore whether influencer content exhibits heterogeneous effects (HTE) across audience segments and test if influencers’ endorsement induces selection consumption, potentially compensating for CGTN's limitations through targeted reach and resonance. Specifically, do influencers perform better within their audiences or achieve breakthrough persuasion and even selection beyond them? Fig. 8 presents heterogeneous treatment effects (HTEs) on objections to the Chinese government's actions in Xinjiang. I focused on four audience segments with stronger theoretical underpinnings: racial groups (White vs. other), prior impressions of China (positive vs. negative), exposure to political influencers (high vs. low), and media trust (high vs. low).[14]

[14] Exposure to political influencers and media trust are not included in the main model as they are not directly related to respondents’ perception and policy positions toward China. See Appendix E for the survey questions.

**Fig. 8. HTE on objections to the Chinese government's actions in Xinjiang**




Note: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

In the top-right panel of Fig. 8, influencers' videos had a strong negative effect (-0.8) among white respondents and a near-zero effect among non-white respondents, providing compelling evidence supporting my theory that ingroup racial identity substantially enhances perceived persuasiveness by reducing foreignness reactance. The remaining three panels explored factors that delineate distinct audience segments and suggest selective consumption patterns. Among respondents with a prior negative impression of the Chinese government, influencers' videos had a substantially stronger effect (-0.7) than CGTN's videos (-0.5). This demonstrates the capacity of influencers' videos to persuade resistant audiences who are also unlikely to consume content from Chinese sources, given their dislike of CGTN's profile, like the White respondents. Influencers' videos also had a stronger effect on high-exposure respondents with low media trust, who are natural audiences for influencers and are less likely to consume news from institutional Chinese sources like CGTN.

While these heterogeneous patterns focus primarily on the persuasive effects of pro-Chinese government videos, communication research has long established a reciprocal relationship between the persuasiveness and credibility of media content and media selection (Shehata & Strömbäck, 2022). This selection effect is particularly intriguing given that the presence of CGTN's profiles reduced the persuasiveness and credibility of the pro-Chinese government content. This is further demonstrated by respondents' answers to the pre-treatment question about their preferred information sources, in which they could choose between one of the influencers' profiles and CGTN's profile. I ran multiple binomial tests to determine if their preference is statistically significant and present the results in Table 1.

**Table 1 Results of Binomial tests – probability of respondents selecting CGTN as an information source about China**

| Sample | Full Sample | Attentive Sample |
|---|---|---|
| Probability of selecting CGNT's profile | 0.444*** (0.154) | 0.455** (0.183) |
| N | 1,044 | 739 |

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 1 confirms this selection dynamic. Respondents were statistically less likely to select CGTN profiles, indicating a consistent preference for influencer profiles across both samples, confirming my earlier hypothesis (**H4**). This suggests a potential self-reinforcing cycle in which influencers' breakthrough capacity converts skepticism toward traditional media and the Chinese government into strategic engagement pathways, potentially transforming one-time avoiders into repeated consumers of propaganda content through self-selected exposure.

**Discussion and Conclusion**

This study advances understanding of foreign propaganda in three keyways. First, it refutes claims that foreign propaganda fails to shape democratic citizens' evaluations of authoritarian regimes—even on contentious human rights issues like Xinjiang (Brady, 2015; Carter & Carter, 2021). Second, it demonstrates credibility as a critical moderator of propaganda's persuasiveness—CGTN's endorsement undermines the persuasiveness of pro-China videos via reduced perceived credibility. Third, it reveals influencers as superior conduits for hard-to-reach audiences. Their content is effective and perceived as more credible than that endorsed by Chinese state media among White respondents, media skeptics, and those with prior negative impressions of the Chinese government. All in all, our findings suggest that influencers serve as a gateway for exposing foreign audiences to pro-autocracy propaganda.

The findings of this paper have broader theoretical and policy implications for understanding the role of online opinion leaders in mediating political communication in an increasingly high-choice information environment (Strömbäck et al., 2024). On the one hand, it showcases how influencers excel at persuading audiences fatigued by formal media through diverse, authentic personal storytelling—consistent with evidence from recent studies that soft, everyday content can better engage authoritarian citizens who are weary of overt propaganda (Lu et al., 2025; Zhu & Fu, 2023). On the other hand, influencer-assisted propaganda poses acute challenges to democratic regulation by rendering foreign influence nearly indistinguishable from organic communication. While systematic labeling of authoritarian-linked entities could disrupt these selection cycles and empower discernment, the fluid, dynamic ecology of influencer networks renders implementation difficult. Moreover, regulating online foreign-backed content risks infringing on free expression, creating a dilemma for democratic governments between free speech and transparency.

This paper also suggests a new pillar of authoritarian resilience: diverse support from foreign citizens. While persuasion through everyday domestic propaganda is an essential pillar of authoritarian stability, this paper suggests that the same logic applies to foreign propaganda of authoritarian regimes. Advances in digital communication enable autocrats to expand propaganda beyond traditional targets such as political and economic elites (Dukalskis, 2021) to ordinary democratic citizens. Although these campaigns may not immediately shift specific foreign policy positions—as evidenced by the null policy effects here—they shape general perceptions of the regime through repeated exposure through official institutions and influencers.

This perceptual shift carries long-term strategic implications for autocratic survival, e.g., lowering foreign and diasporic support for activism against authoritarian regimes (Tsourapas, 2021; Yan & Li, 2023), encouraging foreign investment and trade (Brady, 2015), and exporting authoritarian values to democracies (Mader et al., 2022). By cultivating sympathetic ordinary citizens alongside elites, regimes build diversified resilience against external pressures, with influencers serving as the critical transmission mechanism.

Nevertheless, this study has limitations that future research should address. First, although the selected videos reflect typical Chinese government viewpoints and are matched across sources, unobservable differences beyond narrator identity persist. Future work should employ larger video pools to estimate average treatment effects (ATEs) and standardize content through text-based or AI-generated treatments. Second, the one-time experimental design cannot capture long-term exposure dynamics or influencer-follower interactions. Panel studies tracking follower behavior would reveal how influencers cultivate sustained pro-China attitudes, particularly among initial skeptics induced to follow their profiles. Third, the focus on foreign propaganda targeting democratic audiences overlooks influencers' role in domestic authoritarian contexts, where foreigners serve as neutral outsiders (Fang, 2022). Comparative designs testing foreign versus local influencers across topics would illuminate credibility dynamics in domestic versus cross-border settings. Integrating longitudinal exposure, standardized content, and multi-topic comparisons will strengthen causal identification while mapping influencers' full propaganda potential across audiences.

**Reference**

Abramson, Y., Menon, A., & Gitlin, A. (2025). Whose critique matters? The effects of critic identity and audience on public opinion. *American Journal of Political Science*, *69*(1), 355–370. https://doi.org/10.1111/ajps.12846

Alpermann, B., & Malzer, M. (2024). "In Other News": China's International Media Strategy on Xinjiang—CGTN and New China TV on YouTube. *Modern China*, *50*(2), 135–178. (Sage CA: Los Angeles, CA). https://doi.org/10.1177/00977004231169008

Arı, B., & Sonmez, B. (2025). Human rights violations and public support for sanctions. *Journal of Peace Research*, *62*(1), 68–84. https://doi.org/10.1177/00223433231201450

Barceló, J., & Sheen, G. C.-H. (2024). The Autocracy Bias: Evaluating Democratic Citizens' Perception of Human Rights Violations in Policy Proposals Abroad. *International Studies Quarterly*, *68*(4), sqae135. https://doi.org/10.1093/isq/sqae135

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. https://doi.org/10.1177/0267323118760317

Bestvater, S. (2024). *How U.S. Adults Use TikTok*. Pew Research Center. https://www.pewresearch.org/internet/2024/02/22/how-u-s-adults-use-tiktok/

Blackwell, M., Brown, J. R., Hill, S., Imai, K., & Yamamoto, T. (2025). Priming Bias Versus Post-Treatment Bias in Experimental Designs. *Political Analysis*, *33*(4), 361–377. https://doi.org/10.1017/pan.2025.3

Bradshaw, S., DiResta, R., & Miller, C. (2023). Playing Both Sides: Russian State-Backed Media Coverage of the #BlackLivesMatter Movement. *The International Journal of Press/Politics*, *28*(4), 791–817. https://doi.org/10.1177/19401612221082052

Brady, A.-M. (2015). Authoritarianism Goes Global (II): China's Foreign Propaganda Machine. *Journal of Democracy*, *26*(4), 51–59.

Brockling, M., Hu, H. L., & Fu, K. (2023). The Role of "State Endorsers" in Extending Chinese Propaganda: Evaluating the Reach of Pro-Regime YouTubers. *International Journal of Communication*, *17*, Article 0.

Carter, E. B., & Carter, B. L. (2021). Questioning More: RT, Outward-Facing Propaganda, and the Post-West World Order. *Security Studies*, *30*(1), 49–78. https://doi.org/10.1080/09636412.2021.1885730

Chen, D., & Gao, G. (2023). Chinese Celebrities' Political Signalling on Sina Weibo. *The China Quarterly*, *254*, 466–483. https://doi.org/10.1017/S0305741022001734

Chen, Y., & Yang, D. Y. (2019). The Impact of Media Censorship: 1984 or Brave New World? *American Economic Review*, *109*(6), 2294–2332. https://doi.org/10.1257/aer.20171765

Cheng, Z., Chen, J., Peng, R. X., & Shoenberger, H. (2024). Social media influencers talk about politics: Investigating the role of source factors and PSR in Gen-Z followers' perceived information quality, receptivity and sharing intention. *Journal of Information Technology & Politics*, *21*(2), 117–131. https://doi.org/10.1080/19331681.2023.2173700

Conde, R., & Casais, B. (2023). Micro, macro and mega-influencers on instagram: The power of persuasion via the parasocial relationship. *Journal of Business Research*, *158*, 113708. https://doi.org/10.1016/j.jbusres.2023.113708

Creemers, R. (2017). Cyber China: Upgrading Propaganda, Public Opinion Work and Social Management for the Twenty-First Century. *Journal of Contemporary China*, *26*(103), 85–100. https://doi.org/10.1080/10670564.2016.1206281

Crilley, R., Gillespie, M., Vidgen, B., & Willis, A. (2022). Understanding RT's Audiences: Exposure Not Endorsement for Twitter Followers of Russian State-Sponsored Media. *The International Journal of Press/Politics*, *27*(1), 220–242. https://doi.org/10.1177/1940161220980692

Dai, Y., & Luqiu, L. (2020). Camouflaged propaganda: A survey experiment on political native advertising. *Research & Politics*, *7*(3), 1–10. https://doi.org/10.1177/2053168020935250

Dragojlovic, N. (2015). Listening to Outsiders: The Impact of Messenger Nationality on Transnational Persuasion in the United States. *International Studies Quarterly*, *59*(1), 73–85. https://doi.org/10.1111/isqu.12179

Druckman, J. N. (2001). On the Limits of Framing Effects: Who Can Frame? *The Journal of Politics*, *63*(4), 1041–1066. https://doi.org/10.1111/0022-3816.00100

Dukalskis, A. (2021). *Making the World Safe for Dictatorship*. Oxford University Press. https://doi.org/10.1093/oso/9780197520130.001.0001

Eady, G., Paskhalis, T., Zilinsky, J., Bonneau, R., Nagler, J., & Tucker, J. A. (2023). Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. *Nature Communications*, *14*(1), 62. https://doi.org/10.1038/s41467-022-35576-9

Elswah, M., & Howard, P. N. (2020). "Anything that Causes Chaos": The Organizational Behavior of Russia Today (RT). *Journal of Communication*, *70*(5), 623–645. https://doi.org/10.1093/joc/jqaa027

Escribà-Folch, A., & Wright, J. (2015). *Foreign Pressure and the Politics of Autocratic Survival*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198746997.001.0001

Fan, Y., Pan, J., & Sheng, J. (2024). Strategies of Chinese State Media on Twitter. *Political Communication*, *41*(1), 4–25. https://doi.org/10.1080/10584609.2023.2233911

Fang, K. (2022). Praise from the International Community: How China Uses Foreign Experts to Legitimize Authoritarian Rule. *The China Journal*, *87*(1), 72–91. https://doi.org/10.1086/717550

Geddes, B., & Zaller, J. (1989). Sources of Popular Support for Authoritarian Regimes. *American Journal of Political Science*, *33*(2), 319–347. https://doi.org/10.2307/2111150

Goodwin, A., Joseff, K., Riedl, M. J., Lukito, J., & Woolley, S. (2023). Political Relational Influencers: The Mobilization of Social Media Influencers in the Political Arena. *International Journal of Communication*, *17*, Article 0.

Harff, D. (2022). Political content from virtual 'friends': How influencers arouse young women's political interest via parasocial relationships. *The Journal of Social Media in Society*, *11*(2), Article 2.

Henriksen, F. M., Kristensen, J. B., & Mayerhöffer, E. (2024). Dissemination of RT and Sputnik Content in European Digital Alternative News Environments: Mapping the Influence of Russian State-Backed Media Across Platforms, Topics, and Ideology. *The International Journal of Press/Politics*, *29*(3), 795–818. https://doi.org/10.1177/19401612241230281

Huang, H., & Cruz, N. (2022). Propaganda, Presumed Influence, and Collective Protest. *Political Behavior*, *44*(4), 1789–1812. https://doi.org/10.1007/s11109-021-09683-0

Hudders, L., De Jans, S., & De Veirman, M. (2021). The commercialization of social media stars: A literature review and conceptual framework on the strategic use of social media influencers. *International Journal of Advertising*, *40*(3), 327–375. https://doi.org/10.1080/02650487.2020.1836925

Huddy, L. (2001). From Social to Political Identity: A Critical Examination of Social Identity Theory. *Political Psychology*, *22*(1), 127–156. https://doi.org/10.1111/0162-895X.00230

Huffaker, D. (2010). Dimensions of Leadership and Social Influence in Online Communities. *Human Communication Research*, *36*(4), 593–617. https://doi.org/10.1111/j.1468-2958.2010.01390.x

Imai, K., Jo, B., & Stuart, E. A. (2011). Commentary: Using Potential Outcomes to Understand Causal Mediation Analysis. *Multivariate Behavioral Research*, *46*(5), 861–873. https://doi.org/10.1080/00273171.2011.606743

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334. https://doi.org/10.1037/a0020761

Imai, K., & Yamamoto, T. (2013). Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments. *Political Analysis*, *21*(2), 141–171. https://doi.org/10.1093/pan/mps040

Kane, J. V., & Barabas, J. (2019). No Harm in Checking: Using Factual Manipulation Checks to Assess Attentiveness in Experiments. *American Journal of Political Science*, *63*(1), 234–249. https://doi.org/10.1111/ajps.12396

Keremoğlu, E., & Weidmann, N. B. (2020). How Dictators Control the Internet: A Review Essay. *Comparative Political Studies*, *53*(10–11), 1690–1703. https://doi.org/10.1177/0010414020912278

King, G., Pan, J., & Roberts, M. E. (2013). How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review*, *107*(2), 326–343. https://doi.org/10.1017/S0003055413000014

Kobayashi, T., Zhou ,Yuan, Seki ,Lungta, & and Miura, A. (2025). Autocracies win the minds of the democratic public: How Japanese citizens are persuaded by illiberal narratives propagated by authoritarian regimes. *Democratization*, *0*(0), 1–22. https://doi.org/10.1080/13510347.2025.2475472

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Ladd, J. M. (2010). The Role of Media Distrust in Partisan Voting. *Political Behavior*, *32*(4), 567–585.

Lazarsfeld, P. F., & Katz, E. (1955). *Personal influence: The part played by people in the flow of mass communications*. Free Press.

Liang, F. (2021). *Globalizing Propaganda: Examining the Practice and Regulation of China's Rise on Facebook and Twitter* [Doctoral dissertation, University of Michigan]. https://doi.org/10.7302/2808

Liang, F., Zhu, Q., & Li, G. M. (2023). The Effects of Flagging Propaganda Sources on News Sharing: Quasi-Experimental Evidence from Twitter. *The International Journal of Press/Politics*, *28*(4), 909–928. https://doi.org/10.1177/19401612221086905

Lin, X., Spence, P. R., & Lachlan, K. A. (2016). Social media and credibility indicators: The effect of influence cues. *Computers in Human Behavior*, *63*, 264–271. https://doi.org/10.1016/j.chb.2016.05.002

Loh, D. M. H., & Loke, B. (2024). COVID-19 and the International Politics of Blame: Assessing China's Crisis (Mis)Management Practices. *The China Quarterly*, *257*, 169–185. https://doi.org/10.1017/S0305741023000796

Lu, Y., Pan, J., Xu, X., & Xu, Y. (2025). Decentralized propaganda in the era of digital media: The massive presence of the Chinese state on Douyin. *American Journal of Political Science*, *0*(0). https://doi.org/10.1111/ajps.12990

Mader, M., Marinov, N., & Schoen, H. (2022). Foreign Anti-Mainstream Propaganda and Democratic Publics. *Comparative Political Studies*, *55*(10), 1732–1764. https://doi.org/10.1177/00104140211060277

Mattingly, D. C., & Sundquist, J. (2023). When does public diplomacy work? Evidence from China's "wolf warrior" diplomats. *Political Science Research and Methods*, *11*(4), 921–929. https://doi.org/10.1017/psrm.2022.41

Mattingly, D. C., & Yao, E. (2022). How Soft Propaganda Persuades. *Comparative Political Studies*, *55*(9), 1569–1594. https://doi.org/10.1177/00104140211047403

Mattingly, D., Incerti, T., Ju, C., Moreshead, C., Tanaka, S., & Yamagishi, H. (2024). Chinese state media persuades a global audience that the "China model" is superior: Evidence from a 19-country experiment. *American Journal of Political Science*, *69*(3), 1029–1046. https://doi.org/10.1111/ajps.12887

Mcclain, C., WidJaya, R., Rivero, G., & Smith, A. (2021). *The Behaviors and Attitudes of U.S. Adults on Twitter*. Pew Research Center. https://www.pewresearch.org/internet/2021/11/15/the-behaviors-and-attitudes-of-u-s-adults-on-twitter/

Mena, P., Barbe, D., & Chan-Olmsted, S. (2020). Misinformation on Instagram: The Impact of Trusted Endorsements on Message Credibility. *Social Media + Society*, *6*(2), 2056305120935102. https://doi.org/10.1177/2056305120935102

Messing, S., & Westwood, S. J. (2014). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research*, *41*(8), 1042–1063. https://doi.org/10.1177/0093650212466406

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and Heuristic Approaches to Credibility Evaluation Online. *Journal of Communication*, *60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Meyer, P. (1988). Defining and Measuring Credibility of Newspapers: Developing an Index. *Journalism Quarterly*, *65*(3), 567–574. https://doi.org/10.1177/107769908806500301

Miller, J. M., & Krosnick, J. A. (2000). News Media Impact on the Ingredients of Presidential Evaluations: Politically Knowledgeable Citizens Are Guided by a Trusted Source. *American Journal of Political Science*, *44*(2), 301–315. https://doi.org/10.2307/2669312

Min, B., & Luqiu, L. R. (2021). How Propaganda Techniques Leverage Their Advantages: A Cross-national Study of the Effects of Chinese International Propaganda on the U.S. and South Korean Audiences. *Political Communication*, *38*(3), 305–325. https://doi.org/10.1080/10584609.2020.1763524

Nip, J. Y. M., & Fu, K. (2016). Challenging Official Propaganda? Public Opinion Leaders on Sina Weibo. *The China Quarterly*, *225*, 122–144. https://doi.org/10.1017/S0305741015001654

Nugent, E., & Siegel, A. (2024). How Exiles Mobilize Domestic Dissent. *The Journal of Politics*, *0*(0). https://doi.org/10.1086/734256

Poell, T., Abdulla, R., Rieder, B., Woltering, R., & Zack, L. (2016). Protest leadership in the age of social media. *Information, Communication & Society*, *19*(7), 994–1014. https://doi.org/10.1080/1369118X.2015.1088049

Pornpitakpan, C. (2004). The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Riedl, M. J., Lukito, J., & Woolley, S. C. (2023). Political Influencers on Social Media: An Introduction. *Social Media + Society*, *9*(2), 20563051231177938. https://doi.org/10.1177/20563051231177938

Roberts, M. E. (2018). *Censored: Distraction and diversion inside China's great firewall*. Princeton university press.

Ryan, F., Bogle, A., Ruser, N., Zhang, A., & Impiombato, D. (2021). *Policy Brief: Borrowing mouths to speak on Xinjiang* (No. 55). The Australian Strategic Policy Institute Limited.

Ryan, F., Impiombato, D., & Pai, H.-T. (2022). *Frontier influencers: The new face of China's propaganda | Australian Strategic Policy Institute | ASPI* (No. 65). Australian Strategic Policy Institute. https://www.aspi.org.au/report/frontier-influencers

Schmuck, D., Hirsch, M., Stevic, A., & Matthes, J. (2022). Politics – Simply Explained? How Influencers Affect Youth's Perceived Simplification of Politics, Political Cynicism, and Political Interest. *The International Journal of Press/Politics*, *27*(3), 738–762. https://doi.org/10.1177/19401612221088987

Seitz, A., Tucker, E., & Catalini, M. (2022, March 30). How China's TikTok, Facebook influencers push propaganda. *AP News*. https://apnews.com/article/china-tiktok-facebook-influencers-propaganda-81388bca676c560e02a1b493ea9d6760

Shehata, A., & Strömbäck, J. (2022). Media Use and Societal Perceptions: The Dual Role of Media Trust. *Media and Communication*, *10*(3), 146–157. https://doi.org/10.17645/mac.v10i3.5449

Shen, C., Kasra, M., Pan, W., Bassett, G. A., Malloch, Y., & O'Brien, J. F. (2019). Fake images: The effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. *New Media & Society*, *21*(2), 438–463. https://doi.org/10.1177/1461444818799526

Shmargad, Y. (2022). Twitter Influencers in the 2016 US Congressional Races. *Journal of Political Marketing*, *21*(1), 23–40. https://doi.org/10.1080/15377857.2018.1513385

Strömbäck, J., Lindgren, E., Tsfati, Y., Damstra, A., Vliegenthart, R., Boomgaarden, H., Broda, E., Lebernegg, N., & Galyga, S. (2024). Political Opinion Leaders in High-Choice Information Environments: Are They More Informed Than Others? *Mass Communication and Society*, *27*(5), 949–971. https://doi.org/10.1080/15205436.2023.2281311

Tomz, M. R., & Weeks, J. L. P. (2020). Human Rights and Public Support for War. *The Journal of Politics*, *82*(1), 182–194. https://doi.org/10.1086/705741

Tomz, M., Weeks, J. L. P., & Yarhi-Milo, K. (2020). Public Opinion and Decisions About Military Force in Democracies. *International Organization*, *74*(1), 119–143. https://doi.org/10.1017/S0020818319000341

Treem, J. W., Leonardi, P. M., & van den Hooff, B. (2020). Computer-Mediated Communication in the Age of Communication Visibility. *Journal of Computer-Mediated Communication*, *25*(1), 44–59. https://doi.org/10.1093/jcmc/zmz024

Tsourapas, G. (2021). Global Autocracies: Strategies of Transnational Repression, Legitimation, and Co-Optation in World Politics. *International Studies Review*, *23*(3), 616–644. https://doi.org/10.1093/isr/viaa061

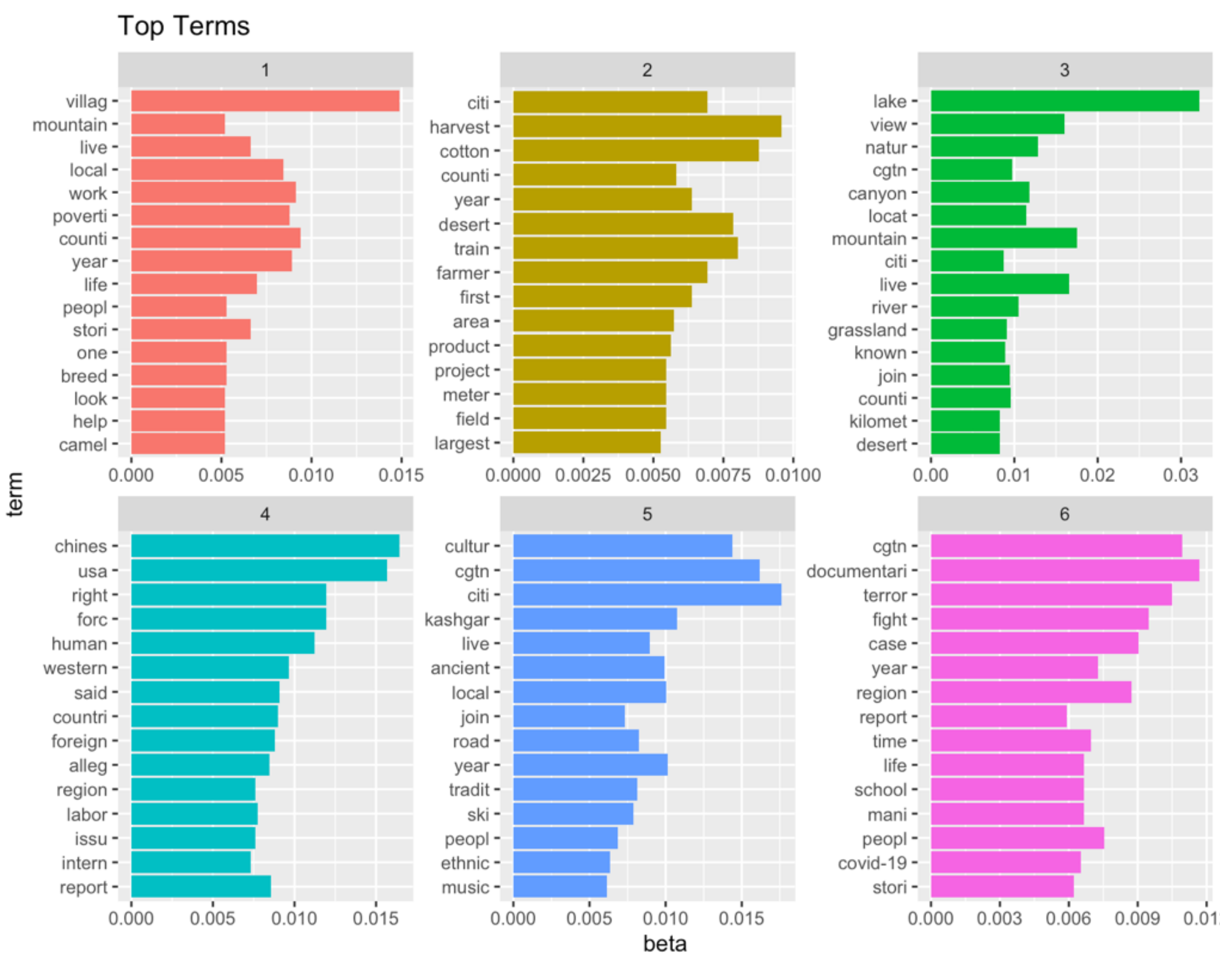
Tucker, J. A., Theocharis, Y., Roberts, M. E., & Barberá, P. (2017). From Liberation to Turmoil: Social Media And Democracy. *Journal of Democracy*, *28*(4), 46–59.
Villegas-Cruz, A. M. (2024a). Nation Branding and Covid-19: An Empirical Investigation of Self-Reports of Medical Donations in Chinese Digital Diplomacy. *Journal of East Asian Studies*, *24*(2), 246–271. https://doi.org/10.1017/jea.2024.13
Villegas-Cruz, A. M. (2024b). The Effect of Economic Ties on Digital Diplomacy. *The Hague Journal of Diplomacy*, *19*, 656–690. https://doi.org/10.1163/1871191x-bja10194
Villegas-Cruz, A. M. (2025). Communicating State Repression to the International Community: A Case Study of How China Frames Its Policies in Xinjiang Online. *Foreign Policy Analysis*, *21*(1), orae028. https://doi.org/10.1093/fpa/orae028
Wang, R., & Zhang, Y. (2025). How political influencers amplified Trump's media-bashing rhetoric on Twitter: From synergistic echoing to strategic avoidance, countering, and retooling. *Journal of Information Technology & Politics*, *0*(0). https://doi.org/10.1080/19331681.2025.2508750
Wasike, B. (2023). I am an Influencer and I Approve This Message! Examining How Political Social Media Influencers Affect Political Interest, Political Trust, Political Efficacy, and Political Participation. *International Journal of Communication*, *17*, Article 0.
Woolley, S. C. (2022). Digital Propaganda: The Power of Influencers. *Journal of Democracy*, *33*(3), 115–129. https://doi.org/10.1353/jod.2022.0027
Xu, W. (2025). Pride and Prejudice: The Dual Effects of "Wolf Warrior Diplomacy" on Domestic and International Audiences. *Journal of Conflict Resolution*, *69*(5), 839–867. https://doi.org/10.1177/00220027241276250
Yan, X., & Li, L. (2023). Propaganda beyond state borders: The deployment of symbolic resources to mobilize political support among the Chinese diaspora. *The Pacific Review*, *36*(3), 433–462. https://doi.org/10.1080/09512748.2021.1968020
Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge university press. https://books.google.com/books?hl=zh-TW&lr=&id=83yNzu6toisC&oi=fnd&pg=PR8&dq=Zaller+1992&ots=6rBAlEXuOT&sig=WxGLeobkK9a2xOjadbddD_toCH8
Zhu, Y., & Fu, K. (2023). How Propaganda Works in the Digital Era: Soft News as a Gateway. *Digital Journalism*, *12*(6), 753–772. https://doi.org/10.1080/21670811.2022.2156366
Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political Effects of the Internet and Social Media. *Annual Review of Economics*, *12*(1), 415–438. https://doi.org/10.1146/annurev-economics-081919-050239

## Appendix A: Treatment Selection

*A1 Analysis of CGTN's YouTube Video about Xinjiang*

Following Mattingly et al. (2024), I select CGTN videos by analyzing CGTN's YouTube content on Xinjiang. I downloaded all videos from CGTN's YouTube Channel containing the keyword "Xinjiang" published between January 1, 2020, and April 30, 2024, totaling 1,335 videos. I run an LDA topic model on this text corpus to identify the topics of CGTN's videos based on high-frequency words, as shown in Fig. A1.

**Fig. A1. High-frequency words from LDA topic models of CGTN's Xinjiang-related YouTube videos.**




Among these topics, Topics 1 and 3 (32.4% of all videos) are about Xinjiang's natural scenery. Topics 2 and 5 (37.0%) are about the more sensitive issues of cotton harvesting and the preservation of traditions and culture in Xinjiang. Finally, Topics 4 and 6 (30.5%) touch upon the most controversial issues of state-led human rights violations in Xinjiang and government policies to fight against terrorism by sending Uyghurs to vocational schools (re-education camps). This evidence suggests that a majority of CGTN's videos about Xinjiang (Topics 2, 4, 5, and 6 – accounting for 67.5% of all videos) are designed to respond to accusations of human rights violations.

The duration of these videos provides another clue to their qualitative differences: videos that touch on Topics 4 and 6 are much shorter (around 8 minutes), while videos covering other topics are longer than 10 minutes and even reach three hours for Topic 3. Thus, I searched for short videos about Xinjiang on CGTN's YouTube using keywords from Topics 2, 4, 5, and 6, and selected three videos corresponding to the three topics mentioned in the main text for further editing.

*A2 Study of pro-Chinese government influencers outside of China*

The pioneering study of Brockling et al. (2023) and other policy research briefs (Ryan et al., 2021, 2022; Seitz et al., 2022) suggest there are more than one hundred Influencers who share their lives and views on China as foreigners using English on X and YouTube, a few dozen of which show a strong pro-Chinese government attitude. I updated this list by searching for keywords such as Xinjiang, Western lies, and Western propaganda on X and YouTube, and by snowball sampling through their social media network, resulting in more than forty accounts on X or YouTube during the data collection period from June to December 2024. Among the foreign influencers I identified, I selected three videos each from three influencers, according to the following criteria:

1. Having constant interaction with Chinese official entities, such as diplomats, state media outlets, and government departments.
2. Being at least a macro-influencer with more than 100,000 followers on either X or YouTube (Conde & Casais, 2023).
3. Showing clues in their social media profiles and videos that they are typical white English speakers with accents from Anglophone or Western European countries.
4. Supporting the Chinese government's position on Xinjiang.

## Appendix B: Additional Data Analysis

*B1 Balance of Data*

Table. A1 and A2 display the demographic balance among experimental groups and between the survey sample and the 2020 U.S. Census. The figure in each cell shows the mean values for the binary demographic traits of gender, college education, age (equal to or above 40), white race, and residence in the U.S. Census Bureau-defined regions: Northeast, Midwest, South, and West. The figures for the whole sample were computed from the survey results, except for residence, which was provided by the survey platform. I used a two-way ANOVA to compare the experimental groups and found they were balanced across most demographic traits, except for residence in the South. The survey sample also largely resembled the U.S. population, with two traits (Aged > 40 and White race) showing differences of ±10%.

**Table A1. Balance of Demographics Data - Experimental Groups (Full Sample)**

| | Control | CGTN's Video | CGTN's Video + Influencers' Profile | Influencer' Video | Influencer' Video + CGTN's Profile | p-value of ANOVA |
|---|---|---|---|---|---|---|
| Male | 0.365 | 0.454 | 0.418 | 0.357 | 0.387 | 0.268 |
| College Education | 0.380 | 0.377 | 0.390 | 0.347 | 0.368 | 0.913 |
| Age > = 40 | 0.740 | 0.743 | 0.714 | 0.745 | 0.753 | 0.894 |
| White | 0.802 | 0.781 | 0.753 | 0.730 | 0.790 | 0.412 |
| Northeast | 0.167 | 0.120 | 0.195 | 0.224 | 0.188 | 0.100 |
| Midwest | 0.286 | 0.191 | 0.244 | 0.240 | 0.231 | 0.313 |
| South | 0.328 | 0.475 | 0.359 | 0.352 | 0.371 | 0.0333* |
| West | 0.219 | 0.213 | 0.202 | 0.184 | 0.210 | 0.927 |

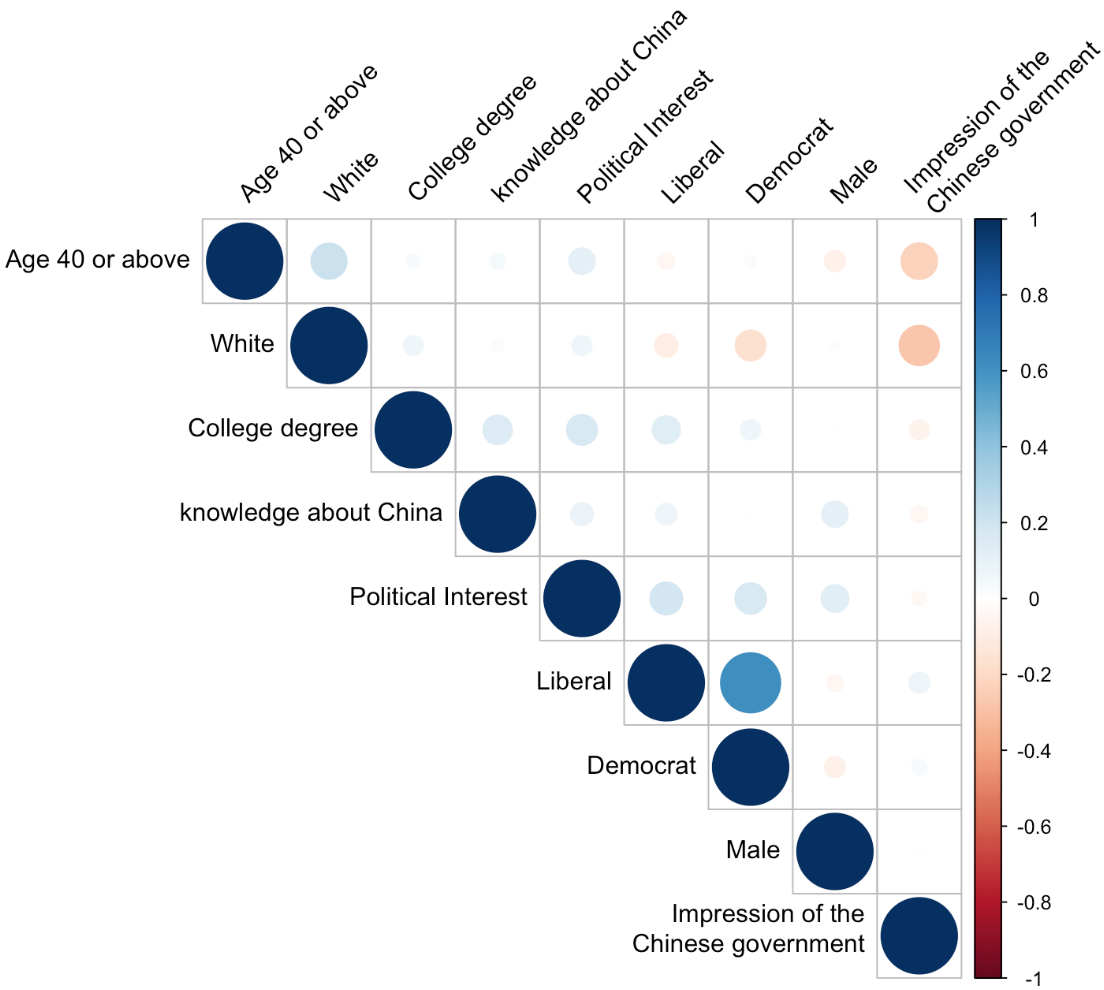
**Table A2. Balance of Demographics Data –**
**Full Sample vs. 2020 U.S Census (Aged 18 or above)**

| | Whole Sample | 2020 U.S Census |
|---|---|---|
| Male | 0.398 | 0.485 |
| College Education | 0.374 | 0.348 |
| Age > = 40 | 0.737 | 0.560 |
| White | 0.769 | 0.641 |
| Northeast | 0.181 | 0.171 |
| Midwest | 0.239 | 0.207 |
| South | 0.375 | 0.351 |
| West | 0.205 | 0.232 |

*B2 Correlation matrix of covariates*

The variables used in our main model show no strong correlation, except between 'democrat' and 'ideology'.

**Fig. A2 Correlation matrix of covariates**



*B3 Tabulated regression results*

**Table A3. ATE - Objections to the Chinese government's actions in Xinjiang**

| | Full sample – unweighted | Full sample – weighted | Attentive sample – unweighted | Attentive sample – weighted |
|---|---|---|---|---|
| *CGTN's Video* | -0.515*** | -0.528*** | -0.514*** | -0.528*** |
| | (0.118) | (0.114) | (0.139) | (0.133) |
| *Influencer's Video* | -0.523*** | -0.526*** | -0.611*** | -0.607*** |
| | (0.116) | (0.112) | (0.136) | (0.131) |
| *CGTN's Profile* | 0.173 | 0.201+ | 0.341* | 0.322* |
| | (0.117) | (0.113) | (0.149) | (0.143) |
| *Influencer's profile* | 0.0749 | 0.0840 | 0.184 | 0.157 |
| | (0.108) | (0.105) | (0.140) | (0.135) |
| Covariate adjusted | No | Yes | No | Yes |
| N | 1,044 | 1,044 | 739 | 739 |

Note: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors in parentheses

**Table A4. ATE - Support for dovish actions against the Chinese government**

| | Full sample – unweighted | Full sample – weighted | Attentive sample – unweighted | Attentive sample – weighted |
|---|---|---|---|---|
| *CGTN's Video* | -0.148 | -0.203+ | -0.221 | -0.241+ |
| | (0.113) | (0.105) | (0.133) | (0.124) |
| *Influencer's Video* | 0.0290 | 0.0167 | -0.0446 | -0.0412 |
| | (0.111) | (0.104) | (0.130) | (0.122) |
| *CGTN's Profile* | -0.0762 | -0.0522 | 0.0915 | 0.0686 |
| | (0.112) | (0.104) | (0.143) | (0.134) |
| *Influencer's profile* | 0.172 | 0.238* | 0.233 | 0.253* |
| | (0.103) | (0.0963) | (0.134) | (0.0454) |
| Covariate adjusted | No | Yes | No | Yes |
| N | 1,044 | 1,044 | 739 | 739 |

Note: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors in parentheses

**Table A5. ATE - Support for Hawkish actions against the Chinese government**

| | Full sample – unweighted | Full sample – weighted | Attentive sample – unweighted | Attentive sample – weighted |
|---|---|---|---|---|
| *CGTN's Video* | -0.0394 | -0.0736 | -0.120 | -0.135 |
| | (0.116) | (0.111) | (0.132) | (0.126) |
| *Influencer's Video* | 0.0592 | 0.0365 | -0.579 | -0.0833 |
| | (0.114) | (0.109) | (0.129) | (0.124) |
| *CGTN's Profile* | -0.0545 | -0.0330 | -0.0417 | 0.00567 |
| | (0.115) | (0.110) | (0.142) | (0.136) |
| *Influencer's profile* | 0.0594 | 0.108 | 0.00952 | 0.0484 |
| | (0.106) | (0.101) | (0.133) | (0.128) |
| Covariate adjusted | No | Yes | No | Yes |
| N | 1,044 | 1,044 | 739 | 739 |

Note: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors in parentheses

**Table A6. ATE - Credibility of the video**

| | Full sample – unweighted | Full sample – weighted | Attentive sample – unweighted | Attentive sample – weighted |
|---|---|---|---|---|
| *Influencer's Video* | 0.0188 | -0.00503 | -0.0129 | -0.0474 |
| | (0.109) | (0.104) | (0.136) | (0.132) |
| *CGTN's Profile* | -0.195+ | -0.215* | -0.283* | -0.274* |
| | (0.108) | (0.103) | (0.133) | (0.128) |
| *Influencer's profile* | -0.0820 | -0.123 | -0.153 | -0.175 |
| | (0.100) | (0.0950) | (0.125) | (0.120) |
| Covariate adjusted | No | Yes | No | Yes |
| N | 852 | 852 | 547 | 547 |

Note: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors in parentheses

*B4 Tabulated results of mediation analysis*

**Table A7. Mediation Analysis**

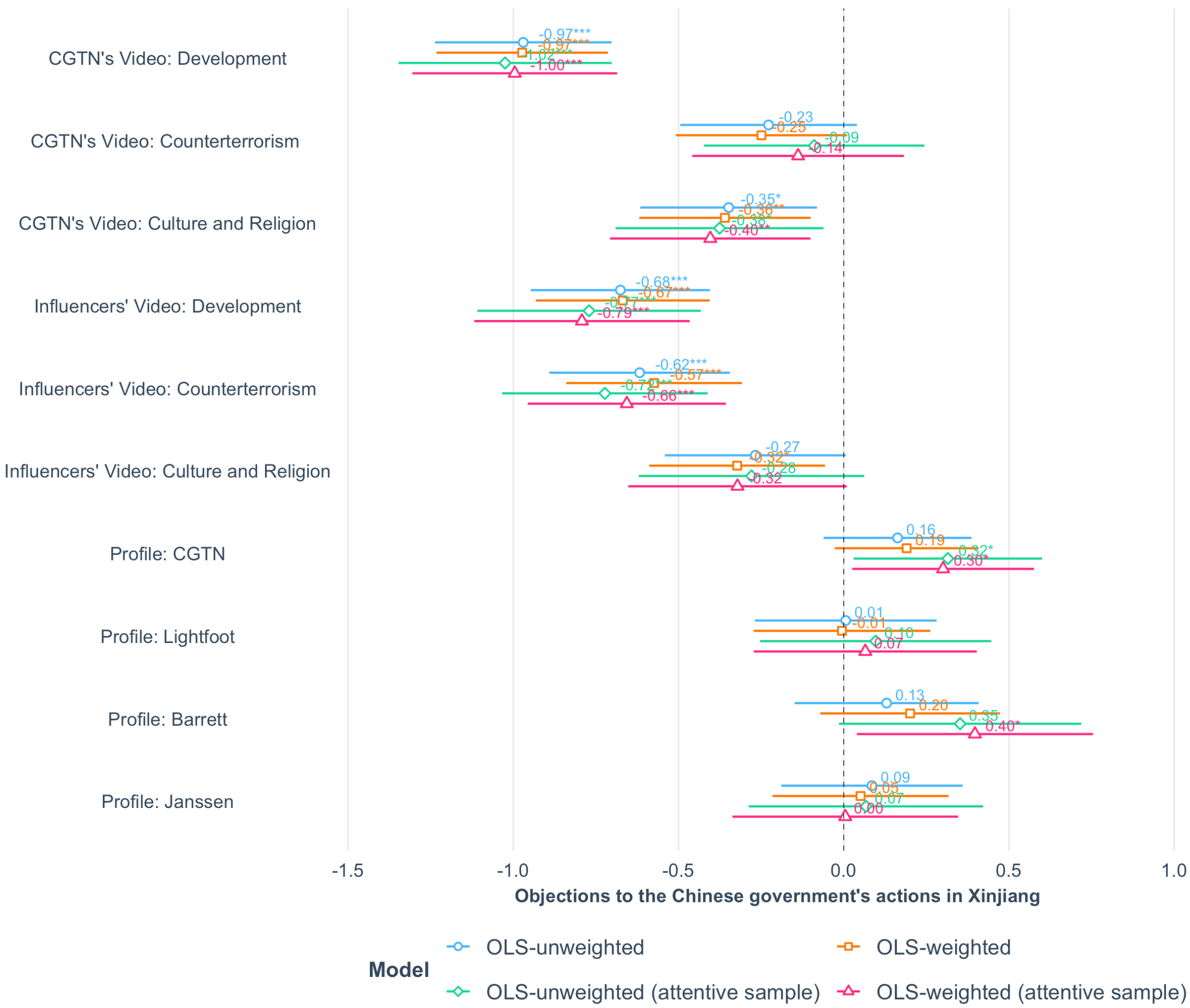
| | Full sample – unweighted | Full sample – weighted | Attentive sample – unweighted | Attentive sample – weighted |
|---|---|---|---|---|
| *ACME* | 0.0983+ | 0.0940+ | 0.162* | 0.141* |
| | [-0.0101, 0.211] | [-0.00773, 0.201] | [0.0151, 0.323] | [0.00236, 0.297] |
| *ADE* | 0.0743 | 0.117 | 0.178 | 0.202 |
| | [-0.124, 0.270] | [-0.0774, 0.314] | [-0.00783, 0.434] | [-0.0468. 0.466] |
| *Total Effects* | 0.173 | 0.211 | 0.340* | 0.343* |
| | [-0.0529, 0.392] | [-0.00584, 0.433] | [0.0403, 0.637] | [0.0543, 0640] |
| Covariate adjusted | No | Yes | No | Yes |
| N | 382 | 382 | 249 | 249 |

Note: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. 95% CI in parentheses

## Appendix C: Robustness Tests

*C1 OLS regression with individual treatment condition*

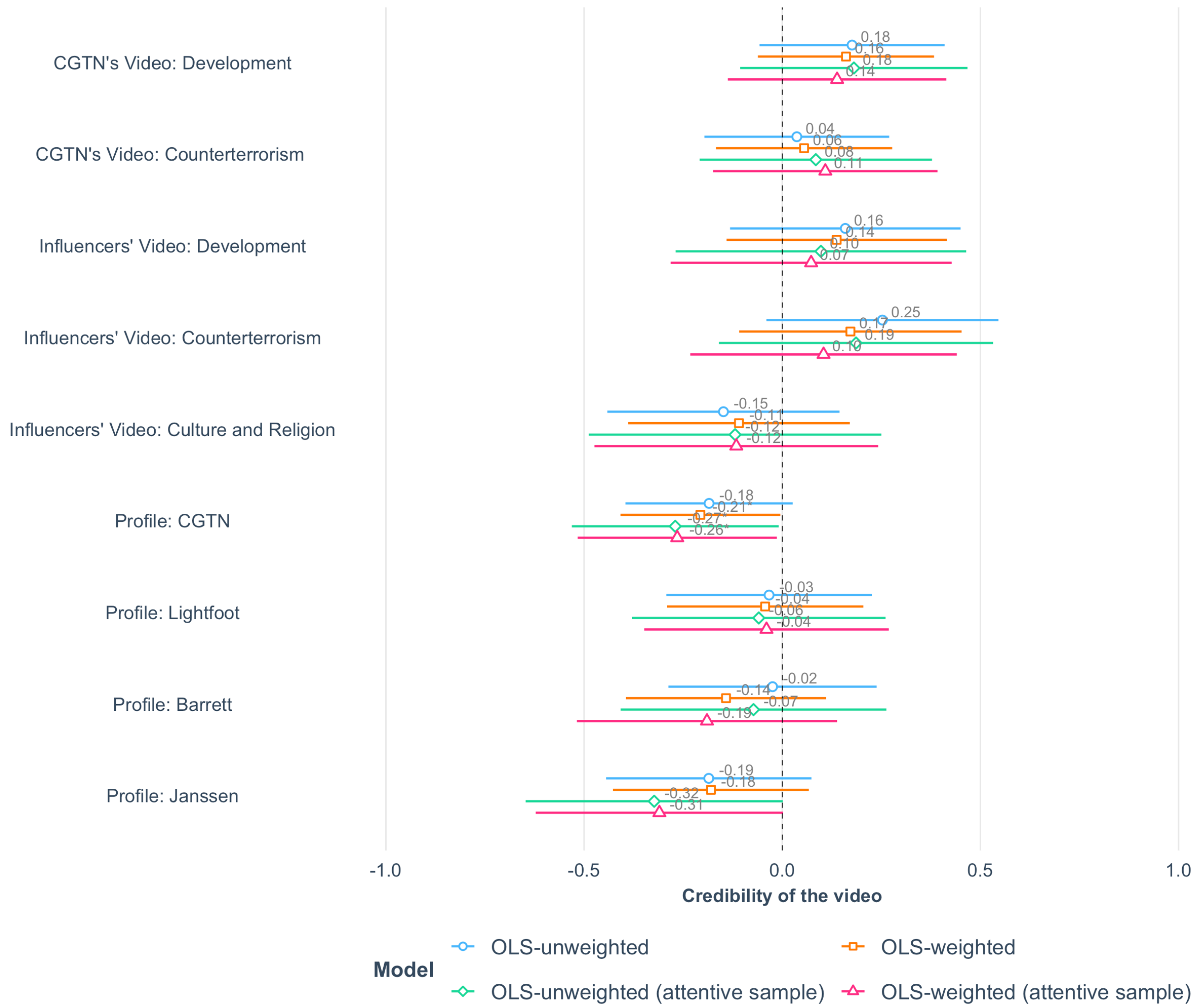
**Fig A3. Persuasive effects of Chinese foreign propaganda by individual treatment condition – objection to the Chinese government's actions in Xinjiang**




Note: * p < 0.05, ** p < 0.01, and *** p < 0.001.

Figure A3 shows that all treatment videos reduced objections to the Chinese government's actions in Xinjiang across models. Development-themed videos produced the largest attitude shifts, as development is core to regime evaluation. Counterterrorism messaging proved effective when delivered by foreign influencers, who positioned security narratives as seemingly independent observations, but ineffective when delivered by CGTN. Cultural and religious content performed relatively stronger when delivered by CGTN than influencers, as audiences expect state media to be better informed on these matters than outsiders. Consistently, CGTN's profile exerted modest negative effects, while influencer profiles had no detectable effect. For credibility, no treatment except CGTN's profile had modest and significant negative effects, which is like that of Fig. 6

**Fig A4. Persuasive effects of Chinese foreign propaganda by individual treatment condition – credibility of the video**




Note: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

*C2 Sensitivity tests of Mediation Analysis*

The robustness of mediation analysis rests on the assumption of sequential ignorability, which consists of two components: (a) treatment assignment is conditionally independent of the outcome and mediator, and (b) the mediator is independent of the outcome given the observed treatment and pre-treatment covariates. In an experimental study, while (a) is always warranted given the random assignment of treatment, (b) could be challenged when the mediator is not randomly assigned and measured after the outcome to avoid a causal effect of the former on the latter via a priming effect, which may, in turn, trigger reverse causality between the two. I address this issue by following Imai et al. (2010) to conduct a sensitivity analysis based on the correlation ($\rho$) of the error terms between the two regression models used in the mediation analysis, which take the outcome and mediator variables as the dependent variables, respectively. In other words, $\rho$ measures the extent to which the results are affected by omitted confounders.

**Fig. A5 Mediation Sensitivity analysis**

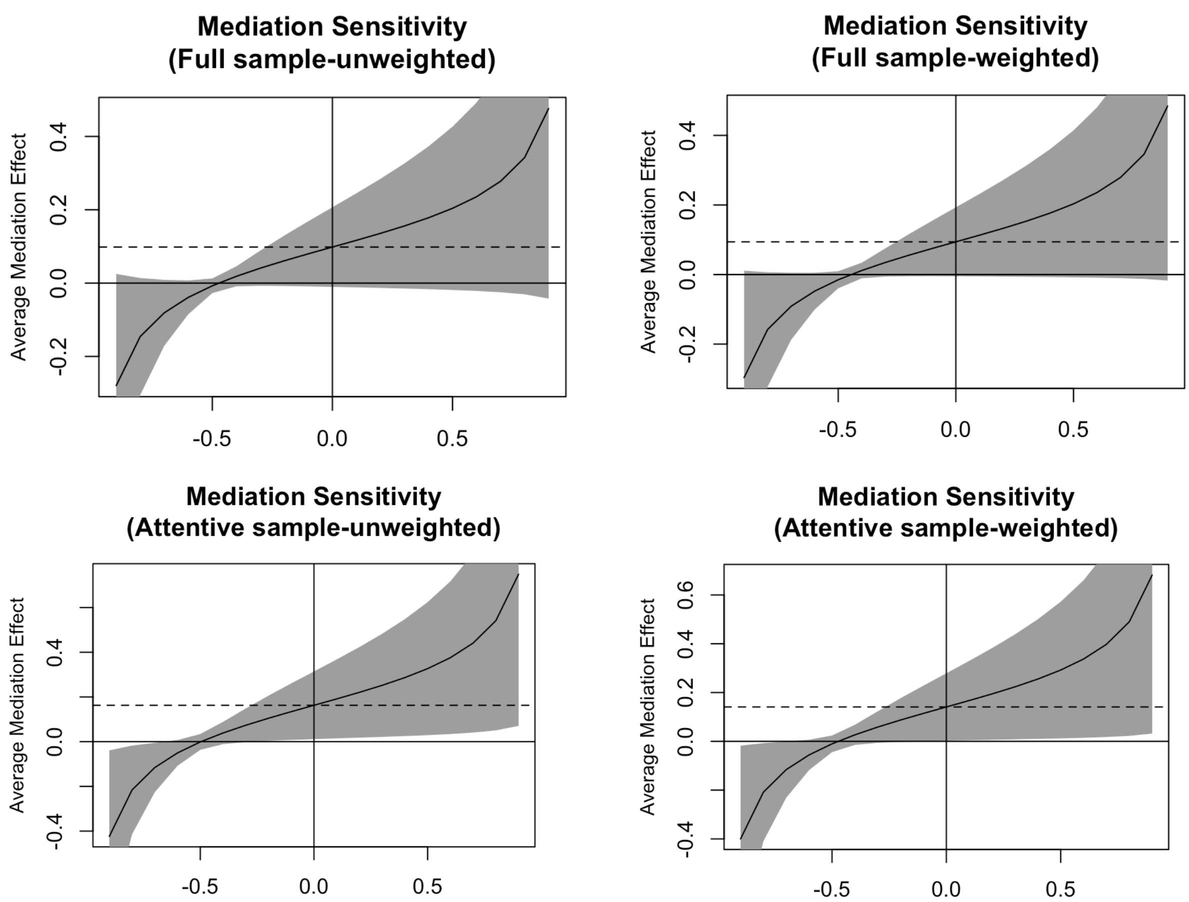


Fig. A5 presents the results, showing the ACMEs and their 95% confidence intervals as a function of $\rho$, ranging from -1 to 1. It shows that, for the mediation analysis on CGTN's profile, its ACME disappears when $\rho$ is around 0.5 and changes sign significantly when $\rho$ is larger than 0.5. While there is no absolute threshold for interpreting the appropriate level of $\rho$, this figure is close to other experimental studies on framing and propaganda (Huang & Cruz, 2022; Imai & Yamamoto, 2013), and such a high correlation of errors in the models is unlikely given all the control variables, suggesting a moderate level of robustness.

## Appendix D: Treatment Videos' Transcripts and Snapshots

*CGTN's Video: Counterterrorism*

[Hui Ao Cui, CGTN's Chinese anchor]: In 2015, like most high school graduates, Dilireba Alimu attended a college entrance exam and was admitted to Shenzhen University. Instead of starting her college life, Dilireba was influenced by her friend and flew to Turkey to attend classes, which taught her the Quran at first, but then something else.

[Dilireba Alimu, Uyghur and ex-reeducation camp inmate, in Mandarin with English subtitles]: That class always taught us heretics are taking out land and natural resources. We (Uyghur people) must fight for what belongs to us and battle for Muslims. They also told us there was a terrorist group that was hiring and providing free training for Muslims to become jihadists.

[Hui Ao Cui, CGTN's anchor]: Two years later, Dilireba returned to Kashir, where she attended her brother's funeral. It was around this time that she heard one of her classmates, Hali Del, had carried out a terrorist attack in Istanbul.

[Dilireba Alimu, Ugyhur and ex-camper, in Mandarin with English subtitles]: My mom was disappointed with me. She persuaded me to study here (Kashi Vocational Education and Training Center).

[Hui Ao Cui, CGTN's anchor]: As part of China's counterterrorism and extremism measures, Xinjiang launched a vocational education and training program. Most of its attendees are influenced by terrorism and extremist thoughts.

[Mijiti Maihemoti, Ugyhur, and Director of Kashi Vocational Education and Training Centre, unknown minority languages dubbed into English]: The people who study here, some of them are suspected of minor criminal offenses, but they do not have to be subject to penalties or criminal punishment. The purpose of establishing the institution is to save and educate them.

**Fig. A6 CGTN's Video: Counterterrorism - video frames every 20 seconds**

*CGTN's Video: Development*
[Hui Ao Cui, CGTN's Chinese anchor]: This elementary school is home to a thousand students. 95% of them are ethnic minorities in China. Everything is free here for students: meals, tuition, and textbooks.

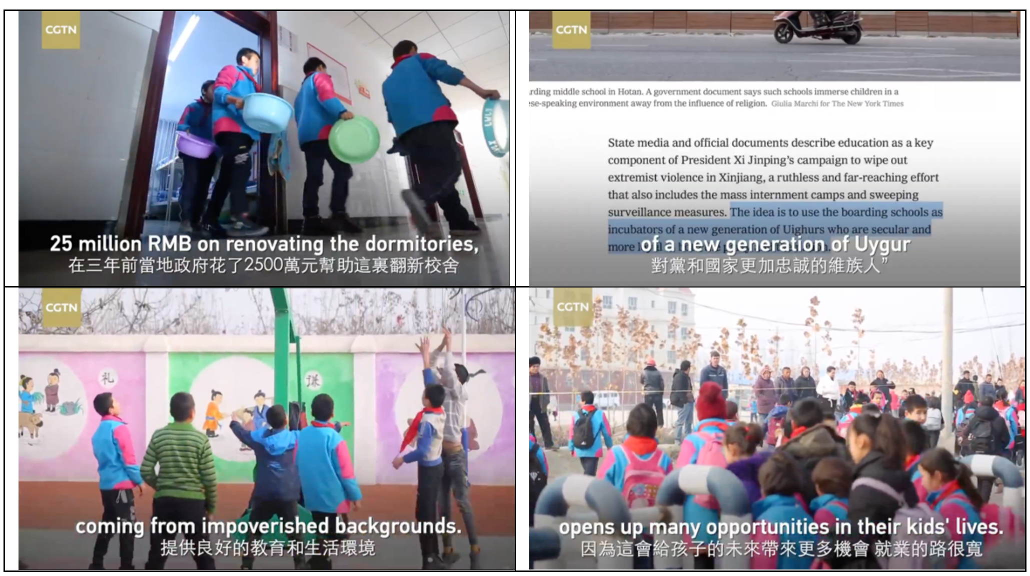
[Hui Ao Cui, CGTN's Chinese anchor]: Three years ago, the local government spent 25 million RMB on renovating the dormitories. That's about 3.6 million U.S. dollars. Additionally, every student receives a subsidy of 1,250 RMB a year

[Hui Ao Cui, CGTN's Chinese anchor]: Unlike in what the New York Times article says that boarding schools serve as incubators of a new generation of Uyghurs who are more loyal to the party and the nation.

[Hui Ao Cui, CGTN's Chinese anchor]: The school's party secretary, Luo Zhihong says, what it's really about is education itself and decent living conditions to children coming from impoverished backgrounds.

[Luo Zhihong, party secretary of a state-sponsored kindergarten, Mandarin dubbed into English]: What we're doing here is solving the education problem for families in poverty, so that their kids don't get left behind. Many Uyghur parents now realize the importance of empowering their kids with knowledge. They support Mandarin education because it opens up many opportunities in their kids.

**Fig. A7 CGTN's Video: Development - video frames every 15-20 seconds**




*CGTN's Video: Culture and Religion*
[Guan Wang, CGTN's Chinese anchor]: Now, looking back, one might think that much of the world has been against China's re-education center in Xinjiang, right?

[Guan Wang, CGTN's Chinese anchor]: Wrong. Yes, 22 Western countries and Japan wrote a letter to the UN criticizing China's Xinjiang "camps. But 54 countries, most of them muslim-majority countries, defended China's counter-extremism efforts in Xinjiang, commending China in its development policies there, and in "providing care to Muslim citizens."

[Guan Wang, CGTN's Chinese anchor]: And they probably have a point. China has built at least 35,000 mosques, more than France's 2,300, America's 2,106, and Great Britain's 1,600. Even on a per capita basis, Chinese Muslims still triple the amount of mosques their Western peers have. That said, there are real concerns too.

[Guan Wang, CGTN's Chinese anchor]: Many reports said that their mosques have been demolished on a large scale. Is that true?

[Mamat Juma, Iman of Kashgar Id Kah Mosque, unknown minority languages dubbed into English]: It is true that some mosques have been demolished in Xinjiang. Those mosques are in disrepair. And they are in danger of collapse. Therefore, they built new mosques after demolishing them. In any village, people can pray in a mosque at any time.

**Fig. A8 CGTN's Video: Culture and Religion - video frames every 15-20 seconds**


*Influencers' Video: Counterterrorism*

[Lee Barett, English White YouTuber]: Xinjiang is a really big area, right? So, which parts of Xinjiang did you go to?

[Noel Lee, Singaporean Chinese YouTuber]: I had gone to Urumqi.

[Lee Barett, English White YouTuber]: Urumqi is the capital, right?

[Noel Lee, Singaporean Chinese YouTuber]: I got to check that one, actually. I would think so because it's a very modernized place. And then Kashgar and Houtan.

[Lee Barett, English White YouTuber]: But do you think generally security is tighter there than in other cities in China? Security was the tightest in Kashgar, followed by Houtan, and then most relaxed in Urumqi. Okay, that's really interesting. Well, why do you think there's that higher level of security?

[Noel Lee, Singaporean Chinese YouTuber]: Well, because Kashgar is the closest to the other countries, right? Neighboring.

[Lee Barett, English White YouTuber]: Ah, yeah, yeah. So, yeah, that's possible because, you know, there's a lot of speculation that there are people returning who do want to cause disruption in Xinjiang. You know, personally, I think that if they have something like that as a threat that they need to deal with it. I mean, I don't know whether you're aware, but previously, two to four or five years ago, there was [sic] a number of terrorists. I have a good friend who's from Xinjiang. And she's told me a number of times that there was a lot of nervousness then. But she says … now when she's in Xinjiang … but when she returns home, she says she feels much more relaxed.

[Noel Lee, Singaporean Chinese YouTuber]: I think we all know about the riots back in the day. But since then till now, there's been no such thing. It's been safe.

**Fig. A9 Influencers' Video: Counterterrorism - video frames every 15-20 seconds**


*Influencers' Video: Development*

[Cyrus Janssen, US White YouTuber]:
Professor Thomas Huber and Professor Helwig Schmidt-Klinze are two German scholars who have been studying and visiting China for over 50 years. After considerable research, including a private investigation on the ground in Xinjiang, the two professors' findings directly contradict many of the central allegations upon which this entire narrative is based.

[Cyrus Janssen, US White YouTuber]:The two professors found that the central government had begun modernizations in education, medical care, and employment. This includes the introduction of 15 years of free education for young Uyghur men and women, including kindergarten, school, and vocational training.

[Cyrus Janssen, US White YouTuber]: In addition, state-subsidized health care has been implemented initially in the southern part of Xinjiang, in addition to free education, with students receiving 200 yuan a month to support their parents.

[Cyrus Janssen, US White YouTuber]: The two China scholars also disclosed that they could not determine a general discrimination against the Uyghur language and culture, but that in Xinjiang, as in all ethnic minority areas.  Mandarin has been established as the main language of instruction in schools starting from the secondary level, while the local Uyghur language is still always offered as a subject in all schools.

**Fig. A10 Influencers' Video: Development - video frames every 15-20 seconds**


*Influencer's Video: Culture and Religion*

[Jason Lightfoot, English White YouTuber]: In Xinjiang, there is [sic] 24,000 mosques in that one single province in China.  If you compare that to the entire country of the United States of America, it is 10 times more.

[Jason Lightfoot, English White YouTuber]: Now you're probably going to say, Oh well, America's not an Islamic country.  Guess what?  Neither is China.  They even have minority scripture on the money, on the actual Chinese national currency.  Wait, I'm going to show you here.  Arabic, you've got Pinyin, Chinese, all different types.  We've got another one here.

[Jason Lightfoot, English White YouTuber]: Is there one single piece of evidence that there's [sic]one million people in all of these camps?  Any evidence?  Where do you get that number from?  It's insane that these huge clothing companies like H&M, Adidas, Nike, … they believe these stories without doing one single bit of research into it themselves.

**Fig. A11 Influencers' Video: Culture and Religion - video frames every 15-20 seconds**


In Xinjiang, there is 24,000 mosques, in
that one single province in China
在新疆，有2.4万所清真寺
Now you're probably going to say, "Oh
well, America is not an Islamic country."
估计有喷子会说，美国不是一伊斯兰国家
大家能看清楚吗？
Any evidence?
这个数据是哪来的？证据在哪里？

**Appendix E: Codebook of variables**

Pre-treatment Variables

1. *Knowledge of Chinese Politics* [High = 1; Low = 0]:
   - High: answer (2) in ChinaQ1 and (2) in China Q2
   - Low: did not answer (2) in ChinaQ1 and (2) in China Q2

ChinaQ1

Which of the following cities is the capital of the People's Republic of China?

- Hong Kong (1)
- Beijing (2)
- Shanghai (3)
- Guangzhou (4)

ChinaQ2

Which of the following persons was NOT the former leader of the People's Republic of China?

- Mao Zedong (1)
- Chiang Kai-shek (2)
- Deng Xiaoping (3)
- Jiang Zemin (4)

2. *Impression of the Chinese government* [Positive = 1; Negative = 0]
   - Positive: answer (3) and (4) in China_Impression
   - Negative: answer (1) and (2) in China_Impression

China_Impression

What is your general impression of the Chinese government?

- Very negative (1)
- Somewhat negative (2)
- Somewhat positive (3)
- Very positive (4)

3. *Profile Selection* [CGTN = 1, Influencer = 0,]
   - Influencer: answer (2), (3), or (4) in CGTN_Preference
   - CGTN: answer (1) in CGTN_Preference

CGTN_Preference

If you were given the following two online sources to seek additional information about China, which of the two would you choose?

- CGTN (1)
- Lee Barrett (2)
- Cyrus Janssen (3)
- Jason lightfoot (4)

4. *Political Interest* [High = 1; Low = 0]
   - High: average of Political_Interest and Political_Attention $\geq 3$
   - Low: average of Political_Interest and Political_Attention $< 3$

Political_Interest
In general, how interested would you say you are in politics?
- Very interested (4)
- Somewhat interested (3)
- Not very interested (2)
- Not interested at all (1)

Political_Attention
How much attention do you pay to news about politics on TV, radio, printed newspapers, or the Internet?
- A great deal (4)
- A lot (3)
- A little (2)
- None at all (1)

5. *Ideology* [Liberals = 1; Non-liberals = 0]
   - Liberal: answered (5) or (4) in Ideology
   - Non-liberals: answered (3), (2), or (1) in Ideology

Ideology
In general, do you think of yourself as…
- Very liberal (5)
- Liberal (4)
- Moderate, middle of the road (3)
- Conservative (2)
- Very conservative (1)

6. *Media Trust* [High = 1; Low = 0]
   - High: answer (4) or (3) in Trust_PreMedia
   - Low: answer (2) or (1) in Trust_PreMedia

Trust_PreMedia
How much, if at all, do you trust the information you get from the National news media, such as ABC, NBC, CBS, etc.?
- A lot (4)
- Some (3)
- Not too much (2)
- None at all (1)

7. *Exposure to political influencers* [High = 1; Low = 0]

- High: average of the items (a) and (b)≥ 3
- Low: average of the items (a) and (b) < 3

Influencer
On social media platforms, there are so-called influencers who have many followers and frequently express their opinions on certain topics. Please indicate how often you engage with them in the following manner.

*Mark one answer in each row*

| | Never (1) | Seldom (2) | Sometimes (3) | Often (4) |
|---|---|---|---|---|
| (a) Randomly come across content from influencers about Politics | | | | |
| (b) Actively search for content from influencers about Politics | | | | |

8. *Instructional Manipulation Check* [Pass = 1; Fail = 0]
   - Pass: answered (3) in Check 1 – included in the main analysis
   - Fail: answered (1), (2), (4), or (5) in Check 1 – filtered in the main analysis

Check1
People have different tastes in movies. For this question, however, we are not interested in your taste but to test whether you can pay attention to instruction, which is important for this research. Please select the option ROMANCE!
- Sci-Fi (1)
- Horror (2)
- Romance (3)
- War (4)
- Comedy (5)

<u>Outcome Variables</u>
1. *Objection to the Chinese government's actions in Xinjiang* [1 – 5]
   - Follow the coding of Poutcome1 in reverse order

Poutcome1
In your opinion, how acceptable or unacceptable are the Chinese government's actions in Xinjiang towards the Uyghurs?
- Very acceptable (5)

- Somewhat acceptable (4)
- Neither acceptable nor unacceptable (3)
- Somewhat unacceptable (2)
- Very unacceptable (1)

2. *Support for Dovish actions* [1 – 5]
   - average of the items (a) and (b) in Poutcome2

3. *Support for Hawkish actions* [1 – 5]
   - average of the items (c) and (d) in Poutcome2

Poutcome2 Would you support or oppose the following U.S. policies against the Chinese government? Mark one answer in each row

| | Strongly support (5) | Somewhat support (4) | Neither support nor oppose (3) | Somewhat oppose (2) | Strongly oppose (1) |
|---|---|---|---|---|---|
| (a) Condemn it publicly | | | | | |
| (b) Impose economic sanctions, even with higher consumer costs in the U.S. | | | | | |
| (c) Threaten military force | | | | | |
| (d) Use military force | | | | | |

4. *Credibility* [1 – 5]
   - Follow the coding of Credibility

Credibility
How much trust and confidence do you have in the video when it comes to showing the situation of Xinjiang and the Uyghurs fully, accurately, and fairly?

- A great deal (5)
- A lot (4)
- A moderate amount (3)
- A little (2)
- None (1)

Post-treatment Variables

1. *Treatment Fact Manipulation Check* [Pass = 1; Fail = 0]
   - Pass: selected (4) in Check3
   - Fail: did not select (4) in Check3

Check3
The video is mainly about ...
*Please select all options you think applicable*
- Education and Development (1)
- (Counter)terrorism and Security (2)
- Religion and Culture (3)
- Xinjiang and the Uyghurs (4)
- Hong Kong and the Hong Konger (5)
- Ukraine and the Ukrainian (6)

2. *Aged above 40* [Aged above 40 = 1; Aged below 40 = 0]
   - Aged above 40: answered (1), or (2) in Age
   - Aged below 40: answered (3), (4), (5), or (6) in Age

Age
How old are you?
- 18 - 29 (1)
- 30 - 39 (2)
- 40 - 49 (3)
- 50 - 59 (4)
- 60 - 69 (5)
- 70 or older (6)

3. *Male* [Male = 1; Female = 0]
   - Male: answered (1) in Gender
   - Female: answered (2) in Gender

Gender
What is your gender?
- Male (1)
- Female (2)

4. *White* [White = 1; Non-white = 0]
   - White: answered (1) in Race
   - Non-white: answered (2), (3), (4), (5), or (6) in Race

Race
Which of the following do you consider to be your primary racial or ethnic group?
- White (1)
- Black or African American (2)

- Latino or Hispanic (3)
- Asian or Asian American (4)
- Native American (5)
- Other (6)

5. *College education* [College degree or above = 1, No college degree = 0 ]
   - College degree or above: answered (4), or (5) in Education
   - No college degree: answered (1), (2), or (3) in Education

Education
What is the highest level of education you have completed?
- Some high school or less (1)
- High school graduate or equivalent (e.g., GED) (2)
- Some college or post high school training (3)
- 4-year college degree (4)
- Postgraduate degree (MA, MBA, MD, JD, PhD, etc.) (5)

6. *Democrats* [Democrats: 1; Non-Democrats: 0]
   - Democrats:
     - answered (1) in Partisan
     - answered (3) in Partisan and (5) in Partisan_Degree
   - Non-Democrats:
     - answered: (2) in Partisan
     - answered (3) in Partisan and (6), or (7) in Partisan_Degree

Partisan
Generally speaking, you think of yourself as a ...
- Democrat (1)
- Republican (2)
- Independent or Other (3)

Partisan_Degree
You think of yourself as ......

Display these choices if Partisan = Democrat
- A strong Democrat (1)
- A not very strong Democrat (2)

Display these choices if Partisan = Republican
- A strong Republican (3)
- A not very strong Republican (4)

Display these choices if Partisan = Independent or Other
- Someone closer to the Democratic Party (5)
- Someone closer to the Republican Party (6)
- Someone neither closer to the Democratic Party nor the Republican Party (7)

**A Replicable Visual LLM Workflow for Emotion Coding in Political Events: Evidence from the 2019 Hong Kong Protests**

Chi Shun (Gary) Fong

PhD Candidate, Department of Political Science, Pennsylvania State University

**Abstract**

Existing research highlights the role of visual elements in triggering emotions such as anger and solidarity, thereby helping overcome the high repression costs in autocracies. However, large-scale studies on visually triggered protest emotions have remained scarce due to methodological challenges in coding these relatively subjective attributes, especially after the rise of social media, which has scaled the problem. The paper proposes using open-weight multimodal Large Language Models (LLMs) with visual analysis capabilities to code protest-related emotions in images. The results show that this LLM approach can achieve near-human-level accuracy in coding social media messages with images from the 2019 Hong Kong protests and can be applied to an alternative dataset of Twitter images related to the Black Lives Matter movement, retrieving a majority of positive cases. Using mid-tier commercial-grade hardware and cloud computing services, these models can typically process thousands of images per day on a single machine. This paper opens the potential of using LLMs to annotate the semantic meanings behind visual elements, which could be applied to study contentious politics and beyond, such as the mobilization effects of images, emotional strategies of protest leaders, framing and misinformation strategies, and themes in online propaganda.

Keywords: social media, protest emotion, visual analysis, large language model

## Introduction

Images play a vital role in political communication. They shape people's political attitudes and guide their behavior by providing information shortcuts for evaluating complex issues (Popkin, 1994) and by eliciting affective responses (Casas & Williams, 2019) that operate beyond narrowly rational calculation (Arceneaux, 2012). Images are particularly important to the study of activism, as visuals are essential to the expression of dissent (Mattoni & Teune, 2014) and help trigger emotions present at every stage of activism (Jasper, 1998). Yet, studies of the visual aspects of activism have remained scarce in sociology and political science, despite the centrality of social movements and contention in these disciplines (Mattoni & Teune, 2014). Most of the existing studies examine how images are utilized for facilitating social movements of various forms and causes, ranging from peaceful protest (Hoffmann & Neumayer, 2025; Vance & Potash, 2022), labor movement (Morrison & Isaac, 2012), environmental activism (Molder et al., 2022), uprising against authoritarianism (Gerbaudo, 2015; Lim, 2013), feminist and gender movement (Palczewski, 2005; Ramsey, 2000; Safaian & Teune, 2022) and far-right movement (Caiani, 2024), and representing these movements in the mainstream media (Arpan et al., 2006; Brown & Harlow, 2019; Corrigall-Brown & Wilkes, 2012; Cowart et al., 2016; Wetzel, 2012) by manually and selectively reading a small number of samples. While these earlier studies provide valuable insights into the potential effects of images on activism, they rarely systematically investigate how images are perceived by citizens *en masse* and how they causally impact their patterns of political participation in observational settings.

The rise of social media partially addresses this issue by providing a systematic collection of visual data with engagement for quantitative analysis, but it also introduces measurement challenges. On the one hand, social media allows users to post multimodal content beyond text, including audio, images, and videos, and provides objective metrics such as likes, shares, reactions, and other engagement indicators to study how visual content is perceived by other users, often mediated by social media algorithms (Neumayer & Rossi, 2018). On the other hand, the abundance of online data requires more efficient and accurate methods for studying image properties. While advances in computational methods like deep learning have enabled thousands of images to be processed within days, the process still entails significant technical barriers and computational costs. Further, these methods focus on image classification, object detection, and face and person analysis (Joo & Steinert-Threlkeld, 2022), leaving the study of more subjective properties, like emotions (Casas & Williams, 2019; Hoffmann & Neumayer, 2025; Jenzen et al., 2020), which are the focus of this paper, largely reliant on qualitative reading of small, selected samples, as in the pre-internet and social media era.

I propose that recent advances in large language models (LLMs) with visual analysis capabilities may offer a more efficient and practical approach to image analysis, even for the relatively subjective task of emotion detection in protest settings. I demonstrate this using three open-weight models—LLaVA (Liu et al., 2023), Gemma3 (Kamath et al., 2025), and Qwen3-VL (Bai et al., 2025)—to analyze protest images related to the 2019 Hong Kong protests. These models represent mainstream medium-sized LLMs that can typically be supported by mid-level commercial-grade hardware or cloud computing services at an acceptable cost, with different advantages given their model architecture.[1] By revising the prompt iteratively based on the model outputs on the results, certainty of the estimation, and reasoning behind the label, the results show that these models can achieve near-human levels accuracy in identifying two major protest emotions, anger and solidarity, particularly the ability to correctly retrieve true cases, across 684 top-viewed human-labeled samples. This approach is validated internally by examining the explanations provided by LLMs and externally by studying another set of human-labeled protest images from the Black Lives Matter movement (Casas & Williams, 2019).

This paper advances the study of the visual dimension of activism by offering a hands-on, efficient, and replicable workflow for using LLMs to code protest images available on social media with near human-level accuracy, with the expectation of future improvement from newer models. By applying state-of-the-art medium-sized open-weight LLMs on mainstream cloud computing services, a single graphics processing unit (GPU), the main hardware that supports LLMs' computation, could typically code at least a few thousand images within a day, at an hourly cost of around 0.13 to 0.45 USD. Researchers could adopt this approach with minimal guidance if they possess basic programming literacy and access to online cloud computing services, without the need to engage in a complicated fine-tuning process. Furthermore, using these models locally instead of calling commercial LLM services via Application Programming

[1] See Appendix D for the comparison of model performance

Interfaces (APIs) makes results highly replicable (Barrie et al., 2024) and reduces the risk of privacy breaches (Balluff et al., 2026). Besides methodological advancement, this paper opens the potential to study substantive questions in contentious politics and beyond that require quantification of semantic meanings behind visual elements, such as the relationship between emotions and protest mobilization, the effectiveness of visual framing on misinformation, and the themes and strategies of propaganda and their persuasiveness.

**Study of visually triggered emotions in contention**

Emotions are broadly defined as socially recognizable patterns of felt response and disposition that transcend individual experiences under a specific societal context (Shah, 2024). Emotions are integral to all social actions. Emotions lead individuals to depart from their normal behavior and remain true to their most deeply held values and attitudes by paying less attention to available information and placing greater reliance on heuristics (Marcus, 2000). As Marcus (2000) noted, Lasswell (2009 [1948]) long ago held that politics is the expression of personal emotions. As such, the study of emotions has been applied across virtually all subfields of political science involving different actors, ranging from issues like voting behavior, partisan polarization, nationalism, and policy process (Marcus, 2000; Shah, 2024; Webster & Albertson, 2022)

Protest participation is no exception (Goodwin et al., 2000; Jasper, 1998; Van Troost et al., 2013). While overlooked by social movement theories that depict protesters as either irrational mobs (Le Bon, 2009) or rationally motivated actors (McCarthy & Zald, 1977), emotions are present at every stage of protest. For instance, prior to a protest, organizers often evoke anger among the public and lead enraged citizens to act against the actor they believe is responsible for their grievance (van Stekelenburg & Klandermans, 2013). After the outbreak of protest, group-based emotions like solidarity and enthusiasm help forge bonds among participants from different backgrounds and sustain their participation (Van Troost et al., 2013). More recent studies in social and political psychology challenge the assumption underlying the distinction between emotion and rationality (Aminzade et al., 2001; Pearlman, 2013; L. E. Young, 2019), demonstrating that emotions alter the utility citizens derive from protest participation by changing their assessment of values, such as security and certainty, perception of reality, risk acceptance, etc.

In this regard, “emotional work” (Hochschild, 1979) – the cultivation of protest-enabling emotions (e.g., anger, solidarity, joy, etc.) and the suppression of protest-disabling emotions (e.g., fear, sadness, etc.) – is a key to organizing successful protests (Gamson, 1990; Jasper, 1998; Pearlman, 2013). Compared with text, visual elements can be perceived universally without prior knowledge (Joo & Steinert-Threlkeld, 2022). Further, visuals can be processed by and are processed more quickly by the brain (Popkin, 1994) and, more importantly, evoke stronger emotional responses than text alone (Casas & Williams, 2019). A large cross-national study of protest-related social media messages demonstrates that images and videos attract more audience engagement than their textual counterparts (Lu & Peng, 2024). In fact, visual elements are widely adopted in protest contexts, either through the explicit efforts of protest leaders or the co-production by online users in the social media era (Neumayer & Rossi, 2018).

While protest may involve more than 20 subtypes of emotions (Jasper, 1998), anger and solidarity are two of the most commonly discussed protest-enabling emotions (Goodwin et al., 2001; Pearlman, 2013)

Anger originates from grievances aroused by perceived injustice – “one is not receiving what one deserves” (Jasper, 1998; van Stekelenburg & Klandermans, 2013), which often involves a responsible out-group, notably institutional actors, as the perpetrator and quest for the restoration of justice (Lambert et al., 2019). Following these studies, I defined anger in this paper as "reactive moral outrage against injustice or perceived violations of rights, typically targeting authorities to frame a clear 'common enemy."[2] It urges citizens to act out of moral outrage despite the risk of repression by making them more risk-averse and optimistic (Pearlman, 2013). Protest organizers often strategically evoke anger through highly visualized symbolic representations of unjustified violence (Olesen, 2015). During the Arab Spring, two well-recognized injustice symbols were Mohamed Bouazizi, a Tunisian street vendor with a college degree who set himself on fire after being repeatedly harassed by the police, and Khaled Said, an

[2] This definition is also employed in the prompt to instruct LLMs, see Appendix A.

Egyptian businessman who was beaten to death by the police after he exposed corruption of the police. Activists circulated their stories and images widely on Arab social media and employed their images as symbols of police brutality and more systemic injustice (Kharroub & Bas, 2016; Lim, 2013; Olesen, 2013). Similarly, in the U.S., activists and social media users spread social media memes and street art targeting police brutality, curating them as symbols of injustice in events like the UC Davis pepper spray incident (Bayerl & Stoynov, 2016) and the Black Lives Matter movement (Casas & Williams, 2019; Vance & Potash, 2022). Across the globe, political parties (Caiani, 2024; Hoffmann & Neumayer, 2025) and civil society organizations (Morrison & Isaac, 2012) frequently employ images to amplify and channel citizens' grievances towards particular targets.

Solidarity, on the other hand, is a group-based emotion that bridges protest participants from different social groups, claims, and preferred tactics (Summers-Effler, 2007) and even turns bystanders into sympathizers (L. Lu et al., 2025). It mobilizes citizens not only through direct persuasion but also by raising the expectation of mass participation (Chwe, 2013; Edmond, 2013). Often, solidarity is nurtured by creating a common identity that allows participants to look past their differences (Neufeld et al., 2019). Based on these attributes, I define solidarity as "positive collective emotion from shared identity, mutual aid, and hopeful commitment that sustains engagement across diverse protest factions."[3] On social media, protesters from the Egyptian revolution, the Spanish indignados, and Occupy Wall Street adopted inclusive, viral, and symbolic protest images as their profile pictures to showcase solidarity with one another, giving rise to a new form of online collective identification (Gerbaudo, 2015). Solidarity is especially important for marginalized groups to consolidate their in-group identification and attract alliances. Gay rights activists in Germany in the 1970s designed posters that encouraged gay men to display their homosexual orientation in public as a symbol of solidarity (Safaian & Teune, 2022). US Suffragists in the 1910s, on the other hand, used cartoons to cultivate solidarity between men and women, both of whom contributed to the war's victory (Ramsey, 2000). Similarly, their counterparts threatened that men would become feminized by women's suffrage in postcards (Palczewski, 2005). Solidarity is also intertwined with and mutually reinforcing of another protest emotion, enthusiasm, which emphasizes pride in the collective

[3] This definition is also employed in the prompt to instruct LLMs, see Appendix A.

identity and hopeful envisioning of the movement (Gerbaudo, 2016; Jasper, 1998). For instance, a study examining Greta Thunberg's Instagram reveals that nearly 80% of her posts invoked a hope narrative, an important component of enthusiasm (Molder et al., 2022).

**Methodological challenges in detecting protest emotions in visual elements**

In contentious politics and social movement studies, research on emotion and visual elements remains scant, despite substantive growth in the past two decades. One simple reason is that the role of emotions in protest was not recognized in earlier social movement theories. While pioneers like Le Bon (2009) framed protest as irrational, impulsive, and driven by a "herd mentality", dominant social movement theories developed from the 1970s, like resource-mobilization (McCarthy & Zald, 1977) and political process (McAdam, 1999), portray protesters as rational players and highlight the strategic deployment of resources. This paradigm posits that emotionality is incompatible with rationality, leaving little space for discussion of the role of emotions in social movements (Goodwin et al., 2000).

However, much of the inattention was also methodological. Measuring affective states accurately and reliably presents a core methodological challenge not only in the study of protest but also in social science in general, as researchers must balance operational scale against psychological depth. To measure individual emotional state at scale, they often employ surveys with structured self-reported instruments like multi-item Likert batteries, providing high precision and construct validity on respondents' conscious affects (Marcus et al., 2017; Watson et al., 1988). However, in protest contexts, it is challenging for researchers to follow the protests using survey methods – with questions ranging from human resources, the spontaneity of protest, identification of protestors, and establishing representativeness (Walgrave et al., 2016; Yuen et al., 2022). Building a representative sample after the protest is also difficult, not to mention concerns about reliability when protesters need to recall their memories (Bradburn et al., 1987) and are susceptible to preference falsification (Kuran, 1991). Occasionally, emotions could be extracted from a comprehensive set of text, such as party manifestos (Crabtree et al., 2020; Jung, 2020; Kosmidis et al., 2019; Widmann, 2021), political speeches (Cochrane et al., 2022; Kosmidis et al., 2019; Osnabrügge et al., 2021), and newspaper articles (L. Young & Soroka, 2012) , notably using dictionary-based lexicons or machine learning algorithms (Grimmer &

Stewart, 2013). For the study of protest emotion, this is less common due to the lack of systematic collection of text that reflects the emotional stage of protesters, though some studies have been successful in achieving this by analyzing discussions on online forums and social media (Shahin & Ng, 2022; Zhu et al., 2022). However, studying text alone ignores the many visual elements used in emotional mobilization. Even worse, relying solely on text may miss how cues operate implicitly to trigger psychological responses. A related famous case is the campaign material used by the Republicans in the 1988 U.S. presidential election that features Willie Horton, an imprisoned NBA player, to signal the stereotyped fear of Black Americans as criminals. Mendelberg's (1997) experimental study on the campaign material found that footage about Horton's case would activate underlying racial prejudice among voters but not the salience of the crime issue, and such effects would disappear when a more explicit cue was presented (Valentino et al., 2002). Such a conclusion would not be reached if these studies solely analyze the text in the campaign material, as both racial and issue cues are presented simultaneously.

For the study of visually triggered protest emotions, researchers manually read protest materials and media representations of protests to infer the intended emotional effects on protesters or the public (Mattoni & Teune, 2014). This approach relies on subjective interpretations and creates observational data points that cannot reliably establish causal connections between emotional and protest-related outcomes, such as success and failure. Further, although manual reading of visual materials provides high-quality, thick descriptions of images, interpretations are also prone to selection bias originating from sampling error.

**Table 1. Comparison of Image Analysis Methods**

| Method | Qualitative Reading | Neural Network Based Models |
|---|---|---|
| **Speed** | Slow | Could be fast, dependent on computational power |
| **Cost** | Generally Low, dependent on sample size | High set-up cost, dependent on availability of cloud computing services |
| **Technicality** | Relatively simple | High |
| **Sample Size** | Relatively small | Large sample, contingent upon computational resource |

| | | |
|---|---|---|
| **Quality** | High but prone to selection and personal biases. | Dependent on the properties to be coded, and quality of human coders if required |
| **Research Focus** | Framing in selected media outlets and protest material | Objective features of images, summary statistics, and audiences' receptivity |

The digital era has made it easier to study visually triggered emotions. Social media platforms are increasingly visual, allowing users to share multimodal content such as images and videos. At the same time, researchers have unprecedented access to these data through platform APIs, web scraping, and commercial data providers, enabling large-scale analyses of visual political communication. As a result, much of the recent research on visually triggered protest emotions relies on social media data rather than traditional print sources (Caiani, 2024; Casas & Williams, 2019; Gerbaudo, 2015; Hoffmann & Neumayer, 2025; Kharroub & Bas, 2016; Lim, 2013; Molder et al., 2022; Olesen, 2013). More importantly, social media provides behavioral measures of audience engagement, such as likes, shares, comments, and reposts, which serve as observable proxies for responses to visual content (Neumayer & Rossi, 2018). For example, Casas and Williams (2019) analyzed Twitter activity surrounding the Black Lives Matter movement and found that tweets containing images—particularly those evoking enthusiasm and fear—were more likely to be retweeted, providing observational evidence that emotionally evocative visuals are associated with greater online political participation.

However, the abundance of available data also creates new challenges. Manually coding hundreds of thousands of images is labor-intensive and often requires large teams of coders, including crowd-sourced workers from platforms such as Amazon Mechanical Turk (MTurk). Without adequate training and quality control, this approach can reduce both the reliability and validity of the resulting data (Lind et al., 2017), particularly if bot or fraudulent accounts are recruited (Wang et al., 2024). For example, Casas and Williams (2019) trained undergraduate research assistants to code only the 1,000 most-retweeted images, while the remaining images were labeled by 1,259 crowdsourced workers. Although a diverse pool of coders may improve the validity of image interpretations, emotional responses to images are inherently subjective, making consistent labeling difficult. Even among the trained research assistants, the intraclass correlation coefficients for emotion coding ranged from 0.19 to 0.54, indicating only low to moderate agreement (Shrout & Fleiss, 1979).

Thus, social scientists increasingly adopt advanced computational methods—notably neural network-based models—to analyze large-scale image datasets. For instance, convolutional neural networks (CNNs), the primary models used in computer vision, recognize local spatial patterns by first converting image pixels into matrices of color intensity. The network then scans these matrices by sliding a filter across the image to extract low-level features, ultimately passing these outputs through successive computational layers to assemble smaller features into complex, large-scale objects (Torres & Cantú, 2022). These models can perform a few common computer vision tasks, including image classification, object detection, and face and person analysis (Joo & Steinert-Threlkeld, 2022). A handful of recent studies apply these advanced models to study protest-related images, such as detecting protest events (Scholz et al., 2025; Sobolev et al., 2020; Steinert et al., 2022; Won et al., 2017; Zhang & Pan, 2019), identifying protest themes (Kim & Bas, 2023; Y. Lu & Peng, 2024; Zhang & Peng, 2024), measuring violence (Rossi et al., 2023; Steinert et al., 2022; Won et al., 2017), and analyzing facial expressions (Prasad & Steinert-Threlkeld, 2025).

**Fig. 1 Examples of images analysis in selected studies**


a) Fig. 5 from Zhang & Pan (2019) – using image and text from social media messages for protest detection


b) Fig.1 from Steinert-Threlkeld et. al (2022) – rating on presence of protest (A), state violence (B) and protester violence (C) based on image features

These studies demonstrate that researchers can process large numbers of images in a reasonable period of time using commercial image analysis services such as Google Vision, Microsoft Face, and Amazon Rekognition, or pre-trained open-source models such as VGGNet (Simonyan & Zisserman, 2015) and ResNet (He et al., 2015). Trained on millions of annotated images, these models can identify thousands of attributes, including persons, objects, scenes, and faces, enabling researchers to generate summary statistics and observe patterns in objective features of visual elements (Peng, Lock, & Ali Salah, 2024). Even when researchers need to analyze images for more specific purposes, such as assessing the level of violence or detecting protest events, they may annotate a small sample of images and use it to update the parameters of existing models, a technique which is called fine-tuning. For instance, Scholz et. al. (2025) employ commercial-grade computers to fine-tune existing models for the detection of protest events from around 140,000 protest-related images. The fine-tuning process took 15 to 30 hours for each model, while the actual labeling process took an hour. Similarly, Zhang & Peng (2024) built their own high-performance computer to categorize 300,000 images using different approaches, with model fine-tuning times ranging from 5 minutes to 24 hours.

These studies face several challenges, however. First, the attributes they label remain objective, relying on features extracted from images. Yet, they can hardly be used to evaluate the semantic meaning of images, the abstract ideological concepts, and the subtleties entailed in a specific social-historical context that are more directly linked to theoretical concepts (Peng, Lock, & Ali Salah, 2024). For instance, a computer vision model can easily detect an everyday object like a yellow umbrella or a three-finger salute from The Hunger Games, but it could not capture how these non-violent elements carry dense symbolic meaning that signals resistance and solidarity in protest movements across Asia (Lertchoosakul, 2021; Veg, 2016). One notable exception is Prasad & Steinert-Threlkeld (2025), who classify facial expressions in protest images as proxies for protest emotions to examine how emotions influence mobilization. However, even this deeper attempt to extract affective states from human faces still falls short of measuring genuine protest emotions. A facial expression, such as a grin, could mean sarcasm or laughing out of anger toward the state, solidarity, or joy toward fellow protesters, depending on context or the actors involved. For instance, a laughing police officer when teargassing protesters, or a politician smiling when responding to protest demands, triggers anger in protesters out of injustice rather than joy.

**Fig.2 Example of images the emotion triggered goes against facial expression in 2019 Hong Kong protests**

a) Hong Kong Police smiled when teargassing protesters


b) Carrie Lam, the Chief Executive (Mayor) of Hong Kong in 2019, responded to journalist during press conference

Second, these approaches focus on images' facial features without referencing the specific context from which these images are drawn. While neural network models can identify visual features, this does not necessarily mean that contextual information can be input into the analysis process. Zhang & Pan (2019), as an exception, use both the message text and the corresponding image to detect protest but treat text and visual analysis as two distinct processes. In our case of detecting protest emotion, analyzing the message text is particularly useful, as automated text analysis is a well-established method of sentiment analysis, especially machine learning methods such as neural network models (Rudkowsky et al., 2018; van Atteveldt et al., 2021). This drawback is more theoretical than methodological. When users scroll past a protest image, carefully crafted textual context anchors, redirects, and/or amplifies their emotional response in ways that cannot be captured solely with the image's pixels. This interplay between text and image is even more critical for analyzing protest, where the text is not an accessory but an integral part of the image itself.

Finally, while online tutorials for applying advanced neural networks are now widely available, some technical barriers remain. First, researchers must possess a solid background in neural network architecture and machine learning algorithms, including selecting appropriate models, ensuring hardware compatibility with software platforms, and fine-tuning and training models for specific tasks. Second, the preprocessing and inference pipelines remain technical and tedious: converting raw social media graphics or posts to standardized image tensors, handling visual noise, and interpreting what the model's layers extract are still "black box" processes. Third, the economic start-up costs pose a steep barrier to entry; locally deploying or fine-tuning these image-processing neural networks requires high-end, dedicated GPU infrastructure that easily exceeds a $3000 USD investment[4], effectively gatekeeping these advanced visual methodologies for resource-constrained researchers, especially those from the developing world.

[4] Taking the set-up of Zhang and Peng (2024) as an example, they used two mid-tier GPU of Nvidia RTX 2080 with a Intel Core i9-9900K CPU. These hardware cost around 2,300 USD in 2022, and similar grade hardware would still cost around 2,000 USD in 2026. Other essential hardware including Motherboard, System Ram, Storage and Power Supply that are compatible with the CPU and GPU would cost at least 1,000 USD

To address these issues, I propose employing open-weight LLMs with visual capabilities to analyze protest images. This approach addresses the key challenges in visual analysis of protest social media. First, LLMs reason through their rich abstraction capabilities for understanding the semantic meanings behind images, especially when the models store large amounts of information relevant to specific cases, like the history of Black civil rights mobilization when studying images related to the Black Lives Matter movement, or the complex history of relations between Hong Kong and Mainland China for analyzing images related to the 2019 Hong Kong protests. These qualities suggest that LLMs could measure more subjective attributes of images, such as emotion, based on information and reasoning stored beyond the pixels (Peng, Lock, & Ali Salah, 2024). Second, some LLMs can integrate information for different modalities, such as text, visual, and audio, in their analysis. Text related to images, such as image captions or social media messages, provides useful context for understanding the semantic meaning behind them, especially for abstract images, thereby improving model performance (Zhang & Leung, 2026).

Finally, the application of LLMs is more hands-on when employing a commercial cloud computing service. The researcher needs only to write proper instructions, called a prompt, to guide the model through the training process, and revise it iteratively by evaluating the results. The steps can be completed online by loading the template code and using a commercial cloud computing service at an affordable price. For example, users could subscribe to Google Colab Pro Plus at a monthly rate of 50 USD. As illustrated in this paper, processing an image would take around 20 to 60 seconds on the platform, and a one-month subscription to the service could purchase enough computational power for analyzing roughly 6,000 – 45,000 images

**Application: Analyzing images from the 2019 Hong Kong protests**

*Overview*

This paper illustrates how to apply open-weight LLMs to analyze images from the 2019 Hong Kong protests to extract protest emotions. I use this case because activists employ both online and offline visual elements for mobilization (Fung, 2020). The messages were originally posted on Telegram - a social media cum instant messaging platform widely adopted by protesters in Hong Kong and in many other authoritarian regimes for its high security and privacy settings

(Herasimenka, 2022; Urman et al., 2021; Wijermars & Lokot, 2022). A core feature of Telegram is the channel, which allows administrators to spread messages to subscribers to facilitate tactical coordination. In early 2020, the author and other researchers at the University of Hong Kong collected around 2 million messages from nearly 1,000 protest-related Telegram channels (Teo & Fu, 2021). For this paper, I focus on a sample from a channel called @hkstandstrong_promo, which posted around 35,000 messages with images during the protests and was the core of the network of channels – it has been mentioned by around 600 channels and once attracted at least 160,000 subscribers. Similar to Casas and Williams (2019), I focus on the most influential messages - 684 out of around 12,000 (~5%) manually coded images with view counts above 30,000. For the emotions, I focus on anger and solidarity, the two well-established protest emotions, as well as Joy. Joy is rarely discussed in the existing literature (Halperin, 2025; Ransan-Cooper et al., 2018), but has the potential to mobilize citizens because, like anger and solidarity, it makes protesters feel relieved, optimistic, and pleased about the movement's progress (Pearlman, 2013). I choose these posts because of their higher interrater reliability, so the evaluation of the LLMs' performance would not be biased by the poor quality or high subjectivity of human coders[5].

*Practical Guide*

<u>Step 1: Prompt Writing</u>

The first step is to write a prompt that instructs LLMs to perform image analysis, specifically to provide binary labels indicating whether each of the three emotions is present (1) or absent (0). I also asked the model to describe the image, provide a short explanation for each coding decision, and rate their certainty in each decision on a scale of 0 to 10. Following some best practices (Giray, 2023), I structured the prompt into three parts: 1) Task; 2) Context; and 3) Output, which was retained despite having rounds of revisions. First, I specify the task as emotion coding of images from the 2019 Hong Kong protests and provide the exact step-by-step instructions. Second, I provide additional context for the coding by listing general rules, definitions, and cues

[5] The human coders are trained undergraduate students, and each image was coded by the agreement of two coders. If disagreement occurs, a third coder is assigned to make the final judgment. The three emotions, anger, solidarity, and joy achieved Fleiss' kappa values of 0.77, 0.84, and 0.62, respectively.

for specific emotions, and examples. Finally, I detailed the model's output schema in JSON format to eliminate output errors.

Step 2: Setting up the LLMs environment

The second step involved setting up the environment to run the local LLMs. Two hardware environments were evaluated: a local high-performance computer and a Google Colab virtual machine with a Pro+ subscription. Rather than relying on API calls to cloud-based LLMs like Gemini and Claude, which pose data privacy risks and suffer from a reliability crisis due to constant updates (Balluff et al., 2026; Barrie et al., 2024), the LLMs were deployed locally. I used Ollama, an open-source framework that enables efficient execution of local models (Marcondes et al., 2025). It supports an extensive library of open-weight LLMs, allowing users to interact with the models locally in interactive notebook environments such as Jupyter and Google Colab. Three medium-sized LLMs with visual analysis capabilities were selected to strike an optimal balance between computational efficiency and task performance (Huang & Wang, 2025), including LLaVA:34B (Liu et al., 2023), Gemma3:27B (Kamath et al., 2025), and Qwen3-VL:32B (Bai et al., 2025).[6]

Step 3: Evaluation and Optimization

The next step is to run the models to analyze the image based on the prompt sequentially and evaluate the results accordingly. Just like the evaluation of standard machine learning tasks, I evaluate the performance of the models on three metrics: 1) Precision, 2) Recall, and 3) F1 scores[7]. While precision measures how accurate a model is when making a positive prediction, recall measures the model's ability to identify all positive cases in the dataset, and F1 is a balanced metric of the two. Then I extract results that do not match the manual labels and read the qualitative description of the image and explanation for each emotion code. In particular, I adopt uncertainty sampling in active learning by focusing on cases the model is least certain about, i.e., those with an uncertainty score close to 0.5 (Rouzegar & Makrehchi, 2024). In a

[6] In the LLM setting, the adjacent "B" indicates the billions of parameters, or variables, that a model uses to process information and make predictions. Larger models generally perform better but require high-tier hardware and consume much more energy. See Appendix D for the comparision of these models

[7] Precision = True Positives/ (True Positives + False Positives), Recall = True Positives/(True Positives + False Negatives), and F1 = 2 x (Precision * Recall)/(Precision + Recall)

traditional supervised machine learning task, these most uncertain cases are labeled by human coders to retrain the model. For LLM inference, I instead updated the prompt by adding new rules and cues for each emotion, while keeping the update as general as possible. This practice helps avoid overfitting, a phenomenon in classical machine learning tasks where a model learns too many nuances rather than general patterns in the current data, leading it to perform poorly on new, unseen data (Giray, 2023).

**Table 2: vLLMs Optimization Strategy by parts**

| Category | Parameter / Variable | Impact on Runtime | Optimization Advice & Guidelines |
|---|---|---|---|
| Model | Parameter Count | **Massive.** Scaling parameters linearly increases compute requirements in Phase 1 and memory bandwidth load in Phase 2. | Use the smallest model that passes the quality threshold. Medium-sized models spanning 20B to 40B provide a good balance on performance and efficiency (Huang & Wang, 2025), especially quantized versions that compute with lower numerical precision (Das et al., 2026). |
| Hardware | GPU Architecture & VRAM Bandwidth | **Massive.** Phase 1 speed scales with GPU’s FLOPS and Phase 2 speed scales directly with VRAM Bandwidth | Set up the system with at least 24 Gigabyte of VRAM that could complete load the model with enough space for the data process, such as a physical Nvidia 3090, or Nvidia L4 or Nvidia A100 on Google Colab. |
| Prompt | Image Resolution | **Moderate:** Higher resolution images are split into larger number of visual patches for input in Phase 1 | Scale down images to the minimum resolution needed, a 768*768 would be an optimal choice – taking up 200 to 800 tokens |
| | Prompt Formatting | **Moderate:** Longer prompts increase Phase 1 time and lead the model to iterate longer in Phase 2 | Enforce direct output instructions and keep the prompt as neat as possible (normally 1 word takes up 1.5 token) |
| Hyperparameters | Context Window Size | **High:** Pre-allocates VRAM for the cache in the computation. Larger | Set it to comfortably fits your input and output size. Avoid |

| | | | |
|---|---|---|---|
| | | size means heavier memory load in Phase 2 | defaulting to maximum window limits |
| | Max Output Limit | **High:** Determines the maximum number of iterations in Phase 2 | Set it as tightly as possible around your expected response length to prevent infinite loop |
| | Temperature | **Moderate:** Determines the amount randomness of the output in Phase 2 | Set it low (0 – 0.2) to reduce the possible variation in the output to reduce the latency and increase the replicability |

Beyond performance, researchers must evaluate computational efficiency when applying LLMs. LLMs' runtime is divided into two distinct computational phases, each governed by different hardware bottlenecks and token parameters—the fundamental units of data processing in LLMs. The first phase processes the input instructions and vision data and is limited by the GPU's floating-point operations per second (FLOPS) and the total length of the input tokens, which, in multimodal contexts, includes textual prompt tokens as well as visual tokens that scale with image resolution and granularity of the LLMs' visual encoders. The second phase generates the output by processing each output sequentially and is bound by the bandwidth of the GPU's memory. Other than this hardware constraint, the efficiency of this second phase is constrained by two factors: the maximum context window, which dictates cache size in GPU VRAM and increases per-token memory retrieval overhead, and the total output token size, which determines the number of iterations required to complete generation. Researchers should use a prompt with optimal size and enough clarity and adjust the maximum context window and total output token size accordingly. Table 3 summarizes how different settings affect LLM runtime.

*Results*

Performance

Table 3 evaluates the performance of three LLMs after three rounds of active learning across two conditions: plain image (evaluating the visual stimulus in isolation) and with message (evaluating the visual stimulus alongside contextual text or message). Performance is measured using Precision, Recall, and F1-score, with baseline ground-truth prevalence rates (T) provided for each emotion.

**Table 3: Performance of the LLMs coding by emotions**

| *Plain image* | LLaVA:34B | | | Gemma 3:27B | | | Qwen3-VL:32B | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1 | Precision | Recall | F1 | Precision | Recall | F1 |
| Anger (T = 31%) | 0.45 | 0.39 | 0.42 | 0.38 | 0.76 | 0.51 | 0.61 | 0.81 | 0.69 |
| Solidarity (T = 65%) | 0.71 | 0.58 | 0.64 | 0.74 | 0.80 | 0.77 | 0.83 | 0.80 | 0.81 |
| Joy (T=14%) | 0.35 | 0.18 | 0.24 | 0.46 | 0.38 | 0.41 | 0.38 | 0.56 | 0.45 |
| *With message* | **LLaVA:34B** | | | **Gemma 3:27B** | | | **Qwen3-VL:32B** | | |
| | Precision | Recall | F1 | Precision | Recall | F1 | Precision | Recall | F1 |
| Anger (T = 31%) | 0.54 | 0.50 | 0.52 | 0.43 | 0.84 | 0.57 | 0.62 | 0.79 | 0.70 |
| Solidarity (T = 65%) | 0.70 | 0.62 | 0.66 | 0.70 | 0.87 | 0.77 | 0.80 | 0.79 | 0.79 |
| Joy (T=14%) | 0.47 | 0.26 | 0.33 | 0.55 | 0.40 | 0.46 | 0.40 | 0.60 | 0.48 |

Overall, model performance varies significantly across emotions, contextual conditions, and underlying architectures, shedding light on the challenges and potential of using LLMs for automated analysis of the semantic meaning of visual elements.

First, providing textual context improves model performance, notably on both LLaVa:34B and Gemma3:27B for labeling Anger and Joy. This improvement was not as pronounced on Qwen3-VL:32B and for the labeling of Solidarity, possibly because there is less room for improvement. Second, the results demonstrate a pronounced performance hierarchy tied to both statistical prevalence and affective complexity. Solidarity—the most prevalent emotion in the benchmark—consistently yields the highest predictive accuracy across all models, peaking at an F1-score of 0.81 with Qwen3-VL:32B. In contrast, Joy (T = 14%) proves substantially harder to detect, topping out at an F1-score of just 0.48. While class imbalance partially accounts for this disparity, there could be a deeper socio-linguistic factor. High-arousal social emotions like Anger and Enthusiasm often manifest through standardized visual gestures and explicit linguistic markers. Conversely, nuanced positive emotions like Joy are heavily context-dependent and culturally mediated, making them far more evasive for LLMs to reliably identify without rich contextual knowledge. The relatively poorer performance could also be attributed to the class imbalance of Joy, which will be further discussed in the validation section below.

Finally, performance gaps exist between models with different architectures[8]. Qwen3-VL:32B emerges as the most robust model, maintaining a strong balance between Precision and Recall for both anger and solidarity, with an F1 score > 0.70, a widely accepted threshold of a good classifier for a balanced dataset[9]. This is because this model is heavily trained on joint-modality analysis, a more fine-grained visual encoder, and more sophisticated reasoning capabilities (Bai et al., 2025). The drawback, as shown in Table 6, is that the model needs a longer runtime to capture more visual details and perform more detailed reasoning. And given the moderate imbalance of the emotion of Anger (T = 31%), having a lower precision and thus an F1 of around 0.50 to 0.60, as in the case of Gemma3:27B, is also considered acceptable and *en par* with recent studies using various machine learning methods to detect emotions in text (Demszky et al., 2020; Koufakou et al., 2024; Mohammad et al., 2018). Based on the results, the performance of LLaVA, which is better adapted to clear photos with obvious subjects, is deemed unacceptable. As emotion is a subjective attribute with ambiguity in human annotation, LLMs' performance is partially bounded by the performance of human coders (Xu et al., 2026). Considering the inter-coder reliability for the three emotions ranged from 0.62 to 0.82, the performance of Qwen3-VL:32B, and to a lesser extent Gemma3:27B, could arguably be regarded as close to human performance when textual data is provided, as in the human coding process.

<u>Reliability and Replicability</u>

---

[8] Refer to Appendix D for the comparison of different models

[9] In machine learning and statistics, a balanced dataset means that the classes or outcomes you are trying to predict are represented in roughly equal proportions.

Beyond predictive performance of individual models, I also explored the agreement between models as a proxy of reliability of results - in other words, are models making similar predictions? As reiterated in the literature, replicability of LLMs matters. So, I explored whether results from local HPC and cloud settings are similar.

**Table 4: Inter-model and Intra-model agreement by emotions**

| *Inter-model (Cohen's Kappa)* | **Anger** | **Solidarity** | **Joy** |
|---|---|---|---|
| plain image | 0.13 | 0.16 | 0.15 |
| with message | 0.22 | 0.18 | 0.25 |
| *Intra-model (Percentage Agreement)* | **Anger** | **Solidarity** | **Joy** |
| LLaVA:34B: plain image | 73% | 69% | 90% |
| LLaVA:34B: with message | 71% | 68% | 89% |
| Gemma3:27B: plain image | 64% | 76% | 85% |
| Gemma3:27B: with message | 68% | 84% | 88% |
| Qwen3-VL:32B: plain image | 83% | 83% | 87% |
| Qwen3-VL:32B: with message | 84% | 80% | 86% |

As shown in Table 4, inter-model agreement, measured via Cohen's Kappa, remains low across all emotions, ranging from 0.13 to 0.16 for plain images and slightly improving to 0.22 to 0.25 when incorporating contextual text messages. This divergence primarily stems from architectural disparities among model families. In Appendix C, I show that Gemma4:31B (Team et al., 2026), which used similar training logic and visual encoding as Qwen3-VL:32B, attained moderate agreement with its results. In terms of intra-model consistency across execution environments, I focused on the percentage of perfect agreement for measuring exact deterministic replicability. All models displayed a higher level of reproducibility, especially for Qwen3-VL:32B (>80% for all emotions). This is especially impressive when the local HPC and Google Colab used different hardware and software configurations, despite using the same size of context windows and output tokens.

Monetary and Time Cost

Table 5 benchmarks the inference runtime per image across three vision-language models under two distinct hardware setups: a local high-performance computing (HPC) environment equipped with an Nvidia RTX PRO 6000 Blackwell GPU (96GB GDDR7), and Google Colab cloud environments (using an Nvidia A100(40GB) for Qwen3-VL:32B, and an Nvidia L4(24GB) for LLaVA:34B and Gemma 3:27B).

**Table 5: Runtimes in second per image by different models and settings**

| | **LLaVA:34B** | **Gemma 3:27B** | **Qwen3-VL:32B**[10] |
|---|---|---|---|
| HPC: plain image | 16s | 11s | 37s |
| HPC: with message | 16s | 11s | 42s |
| Google Colab: plain image | 28s | 22s | 61s |
| Google Colab: with message | 29s | 22s | 73s |

Across both settings, Gemma3:27B was the fastest model, followed by LLaVa:34B, and then the best-performing Qwen3-VL:32B. Considering the performance of the Gemma model in Table 3, it balances performance and efficiency well. Appending textual message prompts created negligible overhead for Gemma 3:27B and LLaVA:34B but added noticeable additional processing time to Qwen3-VL:32 B, reflecting its deeper cross-modal attention mechanisms and reasoning ability. Across all models, the local HPC workstation consistently executed tasks almost 2 times faster than the cloud-hosted Google Colab, given the stronger computational power of the GPU in the HPC. Yet, one should note that GPUs have been severely marked up in recent years given the rising popularity of AI. Hardware capable of running these models can easily exceed 3,000 USD, not to mention that a high-end GPU like the Nvidia RTX PRO 6000 Blackwell in the HPC now costs over 10,000 USD. If the task could be optimized with a budgeted GPU that provides satisfactory performance, using a cloud computing service like Google Colab would be more cost-effective. For instance, a single Nvidia L4 GPU costs only around 0.14 USD per hour and could run nearly 4,000 images per day.

*Validation*

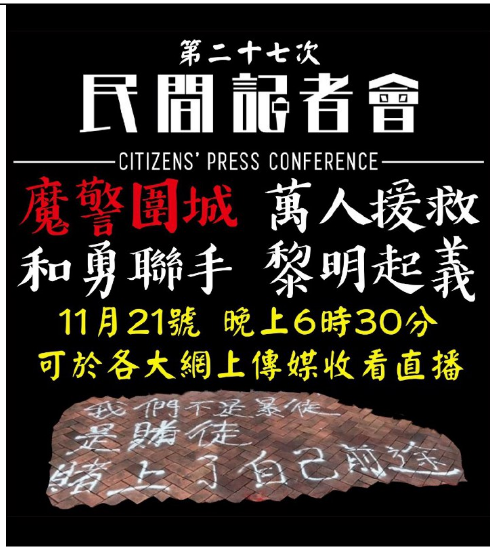
[10] For Qwen3-vl, I use longer context windows of 16,384 and 8,192 for the output token size. For the two models, I use a shorter context window of 4,192 and output token size of 1,024

As stated in the literature (Grimmer & Stewart, 2013), machine learning results must be validated repeatedly. The following session discusses face validity, convergent and divergent validity, construct validity, and external validity, and ecological validity.

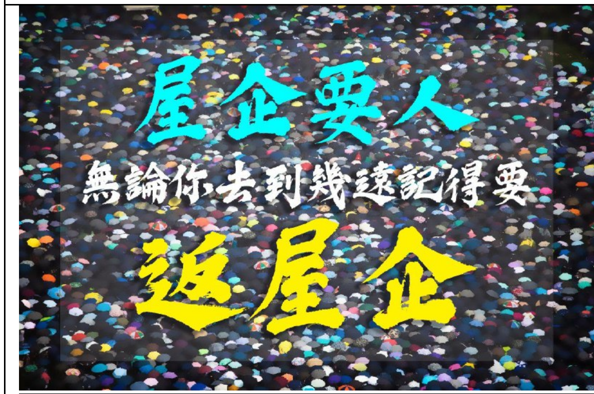
<u>Face Validity</u>

One way to evaluate the content validity of LLM output is to examine the results and assess whether they make sense (Birkenmaier et al., 2024). Thus, I randomly selected cases from the top-performing Qwen3-VL:32B model with text-message context for true-positive, true-negative, false-positive, and false-negative predictions. For the positive cases, they illustrate the model's ability to accurately describe the image, including text that may not be English, and support their decision with proper reasoning. In this regard, LLMs offer a unique advantage over average human coders by enabling applications beyond English-speaking countries. Regarding the negative cases, Qwen3-vl:32B maintains strong reasoning capabilities. Most mispredictions stem from either discrepancies in subjective judgment or subtle contextual knowledge that generally exceeds the scope of a typical large language model (LLM). False positives are primarily due to disagreements; while clues indicating anger and solidarity are present in the images, human coders may judge these cues as insufficiently explicit. Conversely, false negatives are often linked to more definitive details: in the anger image, the eye symbolizes a protester who lost sight due to police gunfire; in the solidarity image, the man is Edward Leung, a notable Hong Kong independence advocate respected by younger activists and jailed before the 2019 protests; and the upper section of the joy image attempts to connect a public indecency incident involving a mainland dancer soliciting tips from elderly citizens in parks with local protests advocating against mainland influence.

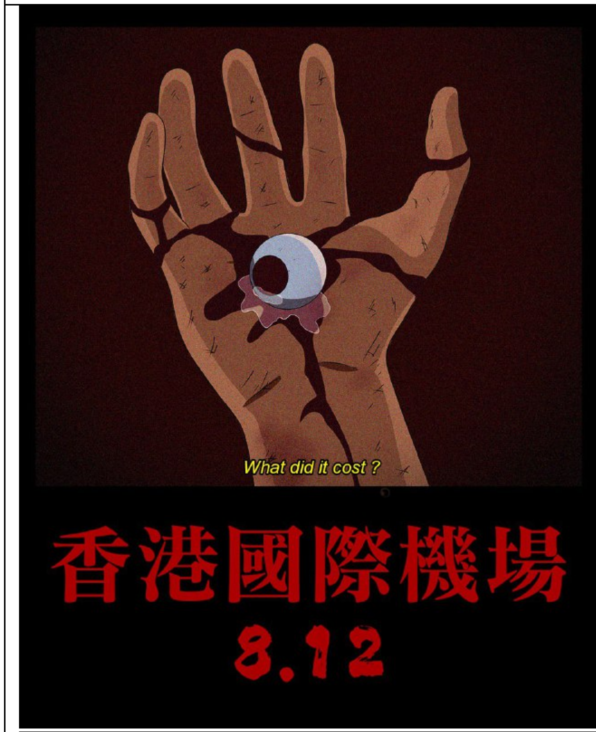
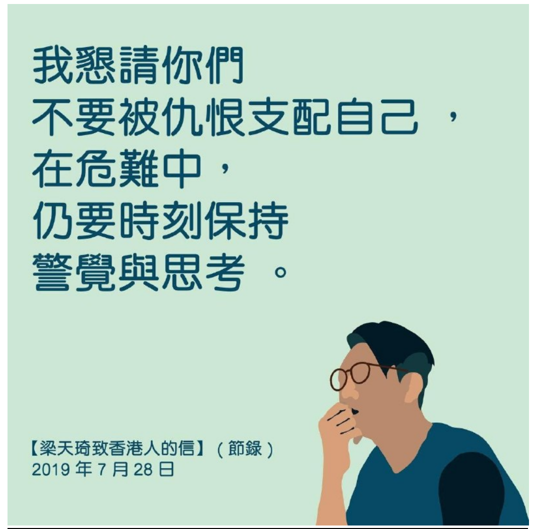
**Fig 3. True positive and true negative prediction – Qwen3-VL:32B**

<table>
<tr><th colspan="3">a) True Positive</th></tr>
<tr><td>Anger</td><td>Solidarity</td><td>Joy</td></tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>Description: Black poster with white, red, and yellow text announcing a press conference, featuring protest slogans and a handwritten note at the bottom.</td>
<td>Description: A large crowd of protesters at night with many holding lights and American flags, overlaid with Chinese text announcing the US Senate passing the Hong Kong Human Rights and Democracy Act.</td>
<td>Description: Blue poster with six yellow icons and text listing practical items for protesters, emphasizing preparation and the five demands, with a P.S. encouraging sharing via Airdrop.</td>
</tr>
<tr>
<td>Reasoning: The image uses the term 'demon police' (魔警) to directly criticize the police, and the context of the event (in response to attacks and blockades by the regime) expresses moral outrage against authorities.</td>
<td>Reasoning: The image depicts a large, unified crowd and highlights international support (US Senate passing the bill), which are key cues for solidarity (shared identity, collective strength, and international backing).</td>
<td>Reasoning: The playful P.S. ("Go crazy and Airdrop...") and lighthearted phrasing (e.g., "Save every dollar, get every dollar") and fun, cartoonish design (with the 'six treasures' as a game-like list) indicate joy and playfulness.</td>
</tr>
</table>

**b) True Negative**

| Anger | Solidarity | Joy |
| --- | --- | --- |
|  |  |  |
| Description: Colorful confetti background with three lines of text in blue, white, and yellow: '屋企要人', '無論你去到幾遠記得要', '返屋企' – Go back home no matters how far you are away | Description: A poster showing a police officer firing a weapon, with text in multiple languages about a 'Hong Kong Police Brutality Database' and details of alleged police misconduct. | Description: A mobile app interface for a protest calendar showing event listings and a December calendar, promoting the app for tracking pro-democracy protests. |
| Reasoning: The image does not contain direct criticism of authorities, documentation of misconduct, or framing of systemic injustice; it is a unifying call without anger. | Reasoning: The image focuses on documenting police brutality and does not explicitly show or reference collective protest unity, mutual aid, or shared identity among protesters. The database is a tool for accountability but lacks direct expressions of solidarity. | Reasoning: The image lacks playful, satirical, or humorous elements; the design is serious and informational, with no evidence of joy. |

| c) False Positive | | |
|---|---|---|
| Anger | Solidarity | Joy |
|  |  |  |
| Description: A large crowd of people in black, many with faces covered, with overlaid text in Chinese and English calling for a protest on October 1st with full gear and unity. | Description: Purple background with two lines of black Chinese text: top line bold and larger, bottom line smaller. Text is in traditional Chinese font. | Description: Light blue background with cartoon of person using laptop. Text in Chinese promotes a bot for submitting propaganda ideas. Lightbulb icon present. |
| Reasoning: The text explicitly references 'fearless of suppression, arrest, terror attack', which is a direct reaction to the authorities' actions (suppression and arrests) and expresses moral outrage against the government and police. | Reasoning: The image is a guide for the 'struggle' (a collective term for the movement) and the post text includes #手足 (comrades), indicating shared identity and mutual aid within the protest community. | Reasoning: The image uses a cartoon style, a lightbulb icon (a playful symbol for an idea), and a lighthearted, conversational tone to encourage participation, which serves as a form of comic relief and morale-boosting. |

<table>
<tr><th colspan="3">d) False Negative</th></tr>
<tr><td>Anger</td><td>Solidarity</td><td>Joy</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
<tr><td>Description: A hand with an eye in the palm, text 'What did it cost?', and below in red: 'Hong Kong International Airport 8.12'.</td><td>Description: Light green background with dark blue Chinese text and an illustration of a man with glasses in a thoughtful pose on the right.</td><td>Description: A protest poster for two 9.21 events: 'Recover Tuen Mun' and 'Yuen Long Sit-in', with a bottom photo of a crowd and the slogan 'Liberate Hong Kong, Revolution of Our Times, Five Demands, Not One Less'.</td></tr>
<tr><td>Reasoning: The image does not contain direct criticism of authorities, documentation of misconduct, or framing of systemic injustice; it is a reflective question without angry cues.</td><td>Reasoning: The text is a cautionary message to avoid hatred and maintain alertness, but it lacks explicit calls for mutual aid, shared identity, or hopeful commitment as defined for solidarity.</td><td>Reasoning: The image lacks any playful, satirical, or lighthearted elements. The top image has some smiling but not in a protest context, and the bottom image is serious. No memes or humor are present.</td></tr>
</table>

Convergent and Divergent Validity

Next, to evaluate whether the LLMs reliably capture distinct affective constructs, I analyze the divergence and convergence of constructs in the model's generated qualitative explanations. By computing cosine similarity between text explanations for different emotions, this tests the conceptual boundaries of the model's classifications, ensuring that its multimodal reasoning maintains clear semantic separation between distinct emotions. I again focused on the best-performing Qwen3-vl:32B model.

**Table 6: Cosine similarity of explanation across emotions (Qwen3-VL32B with message)**

| Emotion Pair Comparison | Mean Cosine Similarity | High Overlap Cases (>0.85) | High Overlap Share (%) |
|---|---|---|---|
| *Solidarity vs. Anger* | 0.5066 | 116 / 684 | 16.96% |
| *Solidarity vs. Joy* | 0.4885 | 116 / 684 | 16.96% |
| *Anger vs. Joy* | 0.5000 | 115 / 684 | 16.81% |

The results confirm that the LLM maintains a healthy conceptual separation, yielding an average explanation cosine similarity of about 0.50 across all emotion pairings. A 0.50 similarity suggests that the model captures the shared domain context of the 2019 Hong Kong protests while maintaining distinct semantic trajectories for each emotion. Furthermore, high-overlap cases were observed in approximately 17% of the dataset. This accurately reflects the data's complexity, as visual elements in protests could allow multiple emotional framing strategies to coexist, such as simultaneous moral outrage against authorities and communal solidarity.

To evaluate the construct validity of the LLMs' output, I apply BERTopic to the model's generated image descriptions grouped by each emotion subset. The rationale is that if the model is truly detecting distinct emotions, the underlying visual motifs extracted should reflect theoretically distinct cues for each emotion. BERTopic accomplishes this by first converting text descriptions into vector embeddings and then clustering semantically similar descriptions. Table 7 again shows the results from Qwen3-VL:32B with message. Table 7 provides robust empirical support for construct validity, demonstrating that the visual pipeline successfully differentiates the thematic content unique to each emotion. Some discovered topics, however, consist of generic background descriptions rather than emotion-specific cues and do not necessarily match the emotions.

**Table 7: BERTopic (Qwen3-VL:32B with message)**

| Emotion Construct | Topic ID | Count | Discovered visual motifs (Keywords) |
|---|---|---|---|
| *Solidarity (N = 442)* | 0 | 35 | Protest demands (demands_protest_airport_protest poster) |
| | 1 | 35 | Rights framing (human rights_rights_poster_human) |
| | 2 | 27 | Uncertain (chinese_black_white_line) |
| | 3 | 24 | Schedules and locations of protests (events_schedule_2019_locations) |
| | 4 | 19 | Uncertain (cartoon_hats_protester_details) |
| *Anger (N = 272)* | 0 | 82 | Police and protester confrontations (police_protesters_officer_showing) |
| | 1 | 18 | Protective gear & protest slogan (mask_gas mask_gas_liberate) |
| | 2 | 17 | Uncertain (chinese_line_person_black) |
| | 3 | 15 | Contextual Markers (kong_hong kong_hong_2019) |
| | 4 | 13 | Criticism (poster_protest poster_protest_criticizing) |
| *Joy (N = 140)* | 0 | 37 | Visual Elements (chinese_police_red_white) |
| | 1 | 20 | General Event Context (2019_protest_hong kong_kong) |
| | 2 | 15 | Cartoons and gear (cartoon_protester_hard hat_hard) |
| | 3 | 12 | Movement claims (demands_protest_poster_protest poster) |
| | 4 | 12 | Assembly site (poster_edinburgh_place_edinburgh place) |

<u>Construct validity</u>

I demonstrate the construct validity of my workflow and that it performs better than similar measurement methods by comparing LLM outputs with those of a facial emotion extraction model and a text-based BERT model.

First, I follow Prasad & Steinert-Threlkeld (2025), who study the relationship between online protest emotion and mobilization, to extract emotional facial expression from the data using the Python package Py-feat (Jin Hyun Cheong et al., 2024) – an open-source package that integrates different machine learning classifiers and neural network models that could be applied to static images and videos. The Py-feat package allows extraction of a range of emotions and related facial attributes, including anger and happiness, which is closely related to Joy. Table 8 below shows the precision, recall, and F-1 score for applying the model to the 684 coded images from the 2019 Hong Kong protests, using different emotion score cut-off thresholds ranging from 1 to 0. The results show that the model could predict anger with moderate precision but not Joy, and that recall for both emotions was extremely low. While the extremely low level of recall is potentially caused by a lack of real human faces in the images included in the samples, the low to moderate level of precision, especially for the more subjective emotion of Joy, suggests the deep semantic meaning behind protest images could not be simply extracted through identification of facial expressions.

**Table 8: Performance of emotional facial extraction model (Py-feat)**

| *Threshold = 0.5* | Precision | Recall | F1 |
|---|---|---|---|
| Anger (T= 31%) | 0.53 | 0.05 | 0.08 |
| Joy (T= 14%) | 0.16 | 0.05 | 0.08 |
| *Threshold = 0.3* | Precision | Recall | F1 |
| Anger (T = 31%) | 0.56 | 0.06 | 0.12 |
| Joy (T= 14%) | 0.14 | 0.08 | 0.11 |

To provide another benchmark for LLMs' performance, I also ran a BERT model (Devlin et al., 2019), specifically the mDeBERTa-v3-base, which supports multilingual analysis, to extract emotions from the text messages in the images. Because images without embedded text were excluded, the effective sample size was reduced to 484 observations. I used an 80/20 train-test split across 5 cross-validation rounds, each with a different random seed. Across the 5 runs, the fine-tuned text-only model excelled at detecting dominant and explicit categories—achieving an average F1-score of 0.83 for Solidarity, outperforming even the best-performing multimodal model, Qwen3-VL with text message (F1 = 0.79). This superior performance likely stems from the fact that solidarity-oriented images frequently contain explicit calls to action and administrative details regarding protest logistics, which are easily captured via text alone. For Anger, mDeBERTa-v3 achieved a solid F1-score of 0.56, though it still fell short of Qwen3-VL with text message (F1 = 0.70). However, the text-only model struggled significantly with Joy (F1 = 0.15, Recall = 0.09), in contrast to Qwen3-VL with text message (F1 = 0.48, recall = 0.6). All in all, LLMs maintained far higher recall and more balanced performance across all three emotion labels. This indicates that zero-shot LLMs possess superior contextual understanding when interpreting subjective visual attributes, particularly given that roughly 30% of the original visual corpus contained no text at all.

**Table 9: Performance of the BERT model on text messages**

| *80% as training data (N= 97)* | Precision | Recall | F1 |
|---|---|---|---|
| Anger (T ~29%) | 0.74 | 0.46 | 0.56 |
| Solidarity (T ~56%) | 0.78 | 0.88 | 0.83 |
| Joy (T ~16%) | 0.56 | 0.09 | 0.15 |

External Validity

For external validation in an alternative context, I apply the above workflow to the Casas & Williams (2019) dataset using the best-performing model, Qwen3-VL:32B. I worked on the top-retweeted 866 images, which were coded by trained undergraduate coders, in addition to MTurk workers who were employed to code. This ensures higher coding quality and, thus, reduces bias in estimating model performance. I focus on the more prevalent and theoretically discussed protest-enabling emotions of anger, sadness, and solidarity-related enthusiasm.[11] As the original dataset is coded on a scale of 0 to 10, I binarized the data into 0 and 1 with a passing threshold of 2, given that the emotional scores were extremely right-skewed and zero-inflated with a mean and median of less than 2. Table 10 presents the results of the model after three rounds of active learning.

[11] I also combined anger and disgust in the data as they are highly theoretically related.

**Table 10: Performance of the LLM coding by emotions – BLM**

| Plain image – Qwen3-VL:32B | Precision | Recall | F1 | Runtime per image |
|---|---|---|---|---|
| Anger (T = 43%) | 0.59 | 0.71 | 0.65 | |
| Sadness (T = 40%) | 0.56 | 0.43 | 0.49 | 34s |
| Enthusiasm (T = 41%) | 0.49 | 0.67 | 0.56 | |
| **With message– Qwen3-VL:32B** | **Precision** | **Recall** | **F1** | **Runtime per image** |
| Anger (T = 43%) | 0.61 | 0.71 | 0.66 | |
| Sadness (T = 40%) | 0.56 | 0.42 | 0.48 | 52s |
| Enthusiasm (T = 41%) | 0.51 | 0.66 | 0.57 | |

As shown in Table 10, the model performance is slightly weaker than the application on the 2019 Hong Kong protests. The classification of Anger and Enthusiasm is marginally acceptable with an F1 of 0.66 and 0.57 when text messages are included in the data. Substantively, high-arousal mobilizing emotions like Anger and Enthusiasm continue to yield stronger overall performance than Sadness, confirming that distinct visual cues remain easier for LLMs to identify than lower-intensity, subtle emotional signals.

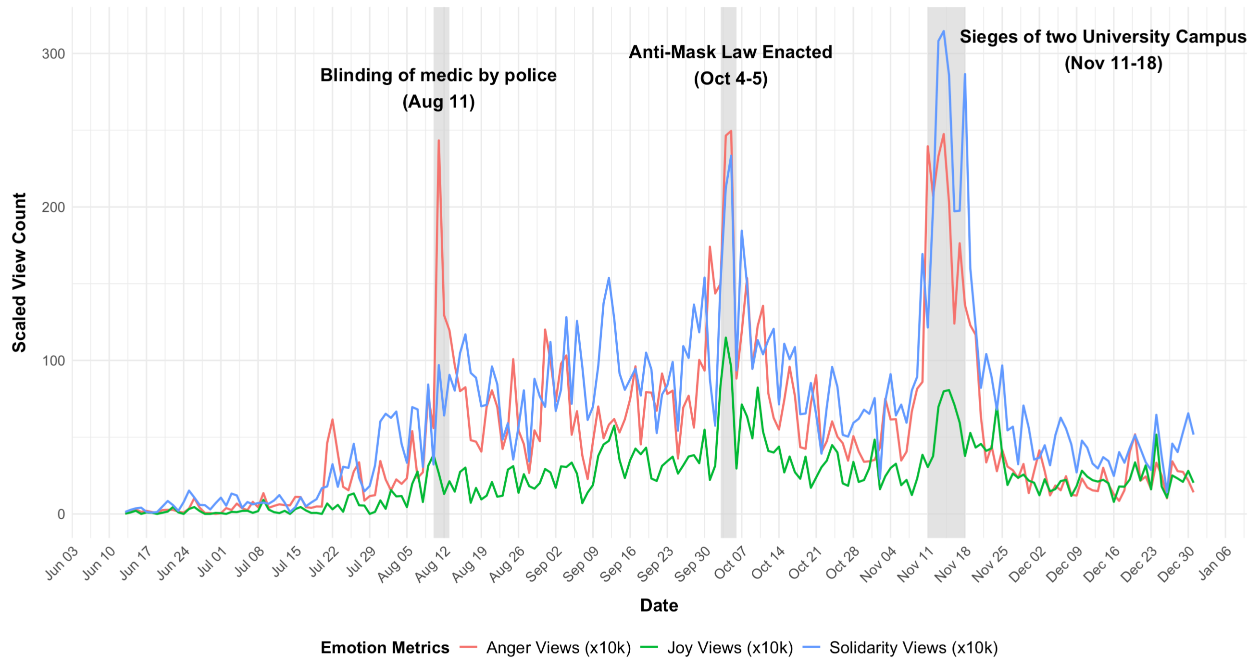
The lower accuracy, however, does not necessarily indicate the inherent incompetence of LLMs in alternative contexts. Instead, the results stem from a combination of platform dynamics, data curation strategies, and human annotation biases. First, the initial Hong Kong protest dataset was drawn from a highly curated, pro-protest Telegram channel specifically designed for political mobilization and "emotional work". These images were explicitly constructed with high signal, emotionally charged iconography, making emotion detection straightforward. By contrast, this application relied on individual Twitter posts from Casas & William (2019) gathered via broad keyword queries with noisy visual signals. This difference in data curation also resulted in a severe zero-inflation of emotions, as they are less prevalent on Twitter than in the specially tasked Telegram channel. From a statistical standpoint, zero-inflation disproportionately penalizes precision by increasing the likelihood of false positives, a component in the denominator of the decision formula. This vulnerability is further compounded by human annotation errors. When relying on crowdsourced workers (e.g., MTurk) for subjective visual tasks, inter-coder reliability tends to be lower. As demonstrated in Monte Carlo simulations by Song et al. (2020), lower inter-coder reliability is more likely to lead to a high false-negative rate. It makes sense as low-effort or uncertain coders rarely assign positive and negative labels with equal probability - systematically defaulting to zero is a safer, harder-to-detect option. Finally, emotional coding is inherently subjective and conditioned by coders' perception of the issue. For a more complex issue with diverse online discussion, like BLM, stronger disagreement over emotional interpretation is possible.

However, there are still ways to improve the model's performance even when the data is noisy, coders' agreement is low, and coding quality is concerning, though the space is limited. First, researchers can leverage the model's self-reported certainty scores to filter out ambiguous predictions and discard low-confidence inferences where visual or textual noise is high to assign a default score, or to continue the active learning process. Second, incorporating a few well-chosen "few-shot" examples (illustrating hard positives and confusing negatives) directly into the prompt text can firmly anchor the model's classification boundaries without requiring any complex architectural fine-tuning. Finally, implementing a lightweight consensus or majority-voting approach by running inference across two models can help smooth out stochastic errors and mitigate the impact of unreliable human baseline annotations.

Ecological Validity

While I have provided evidence of the models' validity across different dimensions, evaluating ecological validity remains necessary. Specifically, does the automated visual emotional coding accurately track the real-world trajectory of the 2019 Hong Kong protests?

**Fig. 4 Emotional Trends Over Time with Key Events in 2019 Hong Kong Protests**




To examine the temporal pattern of visual emotions, I applied the Qwen-VL model to all messages with images posted on the Hong Kong Stand Strong Promote Telegram channel from early June to late December 2019 and calculated the daily aggregated view counts for each emotion. Because the dataset contains over 30,000 images, I treat the message as the unit of analysis and apply the model exclusively to the first image per message. This design assumes that online users' attention focuses primarily on the first image, which also captures the dominant emotional frame. Furthermore, I filter out messages forwarded from external users or channels, as their view metrics reflect reach specifically within this channel rather than across the entire online network, making them systematically lower than those of messages natively posted by this channel.[12] The final dataset contains only 14,354 messages with images. The final analytical sample consists of 14,354 original image-containing messages. Figure 1 displays the resulting temporal patterns of visual emotions alongside major protest milestones.

As shown in Figure 4, the visual emotional dynamics generated by Qwen-VL closely align with critical offline turning points, providing strong ecological validation. First, anger and

[12] Telegram's architecture allows view counts of messages originated from a particular user, group, and channel to be aggregated when displayed in their source. The forwarded messages have an average view count of 407, while the original messages have an average view count of 15,002.

solidarity show high contemporaneity with similar magnitude, while Joy shows a much weaker trend, confirming the dominant role of anger and solidarity in protest emotions. Second, the spikes in emotions correspond to three major milestones in the movement: Anger spiked on Aug 11, matching the blinding of a female first-aider by police projectiles; the enactment of the anti-mask law in early October by the government triggered serious clashes among citizens and mobilization; and the same applies to the sieges of two university campuses by the Hong Kong police in mid-November. Finally, it shows the surge of emotions at the very beginning of the movement and their gradual decay after November, as citizens become increasingly fatigued after a prolonged period of mobilization, especially amid an outbreak of a then-unknown plague of COVID-19. Overall, the tight alignment between LLMs-detected emotional trends and protest patterns confirms that the model's classifications reliably mirror the real-world affective trajectory of the protests.

**Discussion**

This paper illustrates how state-of-the-art open-weight LLMs with visual capabilities could meet the growing demand for multimodal data analysis, driven by the rise of social media, in both accuracy and efficiency. By offering a way to quantify emotions in protests from systematically collected social media data, this study opens a new line of research on the causal relationship between emotions and dissent, a relationship long suggested in the literature but rarely examined beyond experimental settings (Geise et al., 2021; L. E. Young, 2019), and on related questions, such as how protest organizers manage protest emotions alongside state repression. By demonstrating how to annotate semantic meaning in visual elements with LLMs, this paper also unleashes the potential to study beyond contentious politics at scales enabled by visual data. For instance, Peng et al. (2023) suggest going beyond objective features when studying visual misinformation by examining subjective attributes that LLMs can assess, such as image framing and credibility, and the reasoning behind them. Similarly, LLMs could be applied to the study of propaganda, as in Lu et al. (2026), to examine the themes of propaganda videos on Chinese social media beyond small human-annotated samples. This is especially important when popular social media platforms, such as TikTok and Instagram, prioritize visual content over pure text, and memes, short-form video frames, and infographics have become the primary vehicles for political communication(Peng, Lock, & Salah, 2024).

Of course, this paper also has a few limitations. First, the current workflow relies heavily on iterative prompt engineering to optimize inference. Although the results stabilize after 3 rounds of active learning, they may still be sensitive to subtle changes in the input process, such as instruction phrasing, temperature, and context length, which may hinder true out-of-domain generalizability. Researchers with stronger training in machine learning should explore Low-Rank Adaptation (LoRA): fine-tuning LLMs with small, high-quality human-annotated samples can permanently encode domain-specific affective cues directly into weight adapters without compounding runtime costs with a long list of instructions. Second, while active learning via uncertainty sampling refines prompt rules for a specific movement, tailoring rules around edge cases risks overfitting prompt contexts to localized political symbols. Researchers should keep that in mind when revising the prompt. Third, the current workflow focuses on single-model performance and is prone to bias. In the future, researchers could consider formalizing multi-model ensembles to build a more representative decision boundary, improving the reliability and precision of the measurement.

Despite LLMs' ability to analyze visual elements efficiently and with fair accuracy, researchers must still consider several factors when deploying LLMs locally, especially the choice between setting up a higher- performance computer (HPC) and using a cloud computing service.

First, consider the time and monetary costs. Cloud computing services offer a more flexible solution – you can access them as long as you have an internet connection, adjust your usage based on your computational needs, and do not need to maintain HPC day-to-day. By contrast, setting up an HPC requires a certain level of knowledge of hardware and system operation, and a fixed cost that easily exceeds 3,000 dollars. With 3,000 dollars, you can get 30,000 compute units on Google Colab, equivalent to around 20,000 hours of runtime on an Nvidia L4 GPU, or 5,500 hours of runtime on an Nvidia A100 (40GB) GPU. These amounts of runtime should be enough to analyze at least 600,000 to 240,000 images using medium-sized LLMs, assuming each image takes around 30 seconds to process. Moreover, you need not worry about updating the hardware to fit the needs of newer models, as commercial cloud computing services will update their hardware from time to time.

Another concern is replicability. Although LLMs have shown low output variance when running locally with dedicated settings, such as temperature, they remain sensitive to both hardware and software changes. Local HPC provides greater stability in both hardware and software, whereas researchers have no control over the latter when using cloud computing services. In this regard, LLM annotation is like human coding in that exact replication is impossible, but variation should and can be controlled to a certain extent, encouraging researchers to run the data multiple times.

Finally, on the privacy and ethics front—particularly when researching contentious political topics like protests, collective action, and civil disobedience—data sensitivity strongly favors local deployment. Running the model on local HPC, of course, provides stronger data privacy than using a commercial cloud computing service when the image contains identifiable information, such as faces, geolocation data, and metadata, even though the data in this study is publicly available. Nonetheless, the security issue is multilayered; for instance, data from an HPC could be hacked, and LLMs may have backdoors that send data elsewhere unless the device can be kept completely offline, which may complicate daily maintenance. Researchers must assess the risk of a privacy breach and the associated risk when applying LLMs.

**Reference**

Aminzade, R. R., Goldstone, J. A., McAdam, D., Perry, E. J., Sewell, W. H., Tarrow, S., & Tilley, C. (2001). *Silence and Voice in the Study of Contentious Politics* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511815331

Arceneaux, K. (2012). Cognitive Biases and the Strength of Political Arguments. *American Journal of Political Science*, *56*(2), 271–285. https://doi.org/10.1111/j.1540-5907.2011.00573.x

Arpan, L. M., Baker, K., Lee, Y., Jung, T., Lorusso, L., & Smith, J. (2006). News Coverage of Social Protests and the Effects of Photographs and Prior Attitudes. *Mass Communication and Society*, *9*(1), 1–20. https://doi.org/10.1207/s15327825mcs0901_1

Bai, S., Cai, Y., Chen, R., Chen, K., Chen, X., Cheng, Z., Deng, L., Ding, W., Gao, C., Ge, C., Ge, W., Guo, Z., Huang, Q., Huang, J., Huang, F., Hui, B., Jiang, S., Li, Z., Li, M., … Zhu, K. (2025). *Qwen3-VL Technical Report* (arXiv:2511.21631). arXiv. https://doi.org/10.48550/arXiv.2511.21631

Balluff, P., Ho, J. C., Gruber, J. B., Palicki, S., Palmer, A., Rossi, L., Shklovski, I., & Chan, C. (2026). Newer, Larger, Better? A Critique of the Unreflective LLM Adoption in Communication Research. *Political Communication*, *43*(3), 572–581. https://doi.org/10.1080/10584609.2026.2618486

Barrie, C., Palmer, A., & Spirling, A. (2024). *Replication for Language Models*. https://arthurspirling.org/documents/BarriePalmerSpirling_TrustMeBro.pdf

Bayerl, P. S., & Stoynov, L. (2016). Revenge by photoshop: Memefying police acts in the public dialogue about injustice. *New Media & Society*, *18*(6), 1006–1026. https://doi.org/10.1177/1461444814554747

Birkenmaier, L., Lechner, C. M., & Wagner, C. (2024). The Search for Solid Ground in Text as Data: A Systematic Review of Validation Practices and Practical Recommendations for Validation. *Communication Methods and Measures*, *18*(3), 249–277. https://doi.org/10.1080/19312458.2023.2285765

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys. *Science*, *236*(4798), 157–161. https://doi.org/10.1126/science.3563494

Brown, D. K., & Harlow, S. (2019). Protests, Media Coverage, and a Hierarchy of Social Struggle. *The International Journal of Press/Politics*, *24*(4), 508–530. https://doi.org/10.1177/1940161219853517

Caiani, M. (2024). Visual Analysis and the Contentious Politics of the Radical Right. *Sociology Compass*, *18*(9), e13267. https://doi.org/10.1111/soc4.13267

Casas, A., & Williams, N. W. (2019). Images that Matter: Online Protests and the Mobilizing Role of Pictures. *Political Research Quarterly*, *72*(2), 360–375. https://doi.org/10.1177/1065912918786805

Chwe, M. S.-Y. (2013). *Rational ritual: Culture, coordination, and common knowledge* (Reissue with a new afterword by the author). Princeton University Press.

Cochrane, C., Rheault, L., Godbout, J.-F., Whyte, T., Wong, M. W.-C., & Borwein, S. (2022). The Automatic Analysis of Emotion in Political Speech Based on Transcripts. *Political Communication*, *39*(1), 98–121. https://doi.org/10.1080/10584609.2021.1952497

Corrigall-Brown, C., & Wilkes, R. (2012). Picturing Protest: The Visual Framing of Collective Action by First Nations in Canada. *American Behavioral Scientist*, *56*(2), 223–243. https://doi.org/10.1177/0002764211419357

Cowart, H. S., Saunders, L. M., & Blackstone, G. E. (2016). Picture a Protest: Analyzing Media Images Tweeted From Ferguson. *Social Media + Society*, *2*(4), 2056305116674029. https://doi.org/10.1177/2056305116674029

Crabtree, C., Golder, M., Gschwend, T., & Indriđason, I. H. (2020). It Is Not Only What You Say, It Is Also How You Say It: The Strategic Use of Campaign Sentiment. *The Journal of Politics*, *82*(3), 1044–1060. https://doi.org/10.1086/707613

Das, G., La, V., Lau, E., Shrivastava, A., & Gwilliam, M. (2026). *Towards Understanding Best Practices for Quantization of Vision-Language Models* (arXiv:2601.15287). arXiv. https://doi.org/10.48550/arXiv.2601.15287

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A Dataset of Fine-Grained Emotions. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 4040–4054). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.372

Edmond, C. (2013). Information Manipulation, Coordination, and Regime Change*. *The Review of Economic Studies*, *80*(4), 1422–1458. https://doi.org/10.1093/restud/rdt020

Fung, W. L. C. (2020). *Canvas of dissent: A study of visuals and its significance in group formations and communications during the 2019 Hong Kong Anti-ELAB movement* [Master Thesis, Lund University]. https://lup.lub.lu.se/student-papers/search/publication/9009383

Gamson, W. A. (1990). *The strategy of social protest* (2. ed). Wadsworth Publ. Co.

Geise, S., Panke, D., & Heck, A. (2021). Still Images—Moving People? How Media Images of Protest Issues and Movements Influence Participatory Intentions. *The International Journal of Press/Politics*, *26*(1), 92–118. https://doi.org/10.1177/1940161220968534

Gerbaudo, P. (2015). Protest avatars as memetic signifiers: Political profile pictures and the construction of collective identity on social media in the 2011 protest wave. *Information, Communication & Society*, *18*(8), 916–929. https://doi.org/10.1080/1369118X.2015.1043316

Gerbaudo, P. (2016). Constructing Public Space| Rousing the Facebook Crowd: Digital Enthusiasm and Emotional Contagion in the 2011 Protests in Egypt and Spain. *International Journal of Communication*, *10*(0), Article 0.

Giray, L. (2023). Prompt Engineering with ChatGPT: A Guide for Academic Writers. *Annals of Biomedical Engineering*, *51*(12), 2629–2633. https://doi.org/10.1007/s10439-023-03272-4

Goodwin, J., Jasper, J. M., & Polletta, F. (2001). *Passionate Politics: Emotions and Social Movements*. University of Chicago Press.

Goodwin, J., Jasper, J., & Polletta, F. (2000). *The Return of The Repressed: The Fall and Rise of Emotions in Social Movement Theory*. https://doi.org/10.17813/maiq.5.1.74u39102m107g748

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Halperin, L. (2025). Combining emotions: Hope, anger, joy, and love in Israeli peace movements. *Social Movement Studies*, *24*(3), 379–394. https://doi.org/10.1080/14742837.2022.2155628

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (arXiv:1512.03385). arXiv. https://doi.org/10.48550/arXiv.1512.03385

Herasimenka, A. (2022). Movement Leadership and Messaging Platforms in Preemptive Repressive Settings: Telegram and the Navalny Movement in Russia. *Social Media + Society*, *8*(3), 20563051221123038. https://doi.org/10.1177/20563051221123038

Hoffmann, M., & Neumayer, C. (2025). Images of protest in movement parties' social media communication. *New Media & Society*, *27*(8), 4929–4952. https://doi.org/10.1177/14614448241243352

Huang, D., & Wang, Z. (2025). LLMs at the Edge: Performance and Efficiency Evaluation with Ollama on Diverse Hardware. *2025 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN64981.2025.11228317

Jasper, J. M. (1998). The Emotions of Protest: Affective and Reactive Emotions In and Around Social Movements. *Sociological Forum*, *13*(3), 397–424. https://doi.org/10.1023/A:1022175308081

Jenzen, O., Erhart, I., Eslen-Ziya, H., Korkut, U., & McGarry, A. (2020). The symbol of social media in contemporary protest: Twitter and the Gezi Park movement. *Convergence*, 1354856520933747. https://doi.org/10.1177/1354856520933747

Joo, J., & Steinert-Threlkeld, Z. (2022). Image as Data: Automated Content Analysis for Visual Presentations of Political Actors and Events. *Computational Communication Research*, *4*(1), Article 1.

Jung, J.-H. (2020). The Mobilizing Effect of Parties' Moral Rhetoric. *American Journal of Political Science*, *64*(2), 341–355. https://doi.org/10.1111/ajps.12476

Kamath, A., Pathak, S., Vieillard, N., Merhej, R., Perrin, S., Matejovicova, T., Ramé, A., Rivière, M., Rouillard, L., Mesnard, T., Cideron, G., Grill, J., Ramos, S., Yvinec, E., Casbon, M., Pot, E., Penchev, I., Liu, G., Visin, F., … Hussenot, L. (2025, March 25). *Gemma 3 Technical Report*. arXiv.Org. https://arxiv.org/abs/2503.19786v1

Kharroub, T., & Bas, O. (2016). Social media and protests: An examination of Twitter images of the 2011 Egyptian revolution. *New Media & Society*, *18*(9), 1973–1992. https://doi.org/10.1177/1461444815571914

Kim, M., & Bas, O. (2023). Seeing the Black Lives Matter Movement Through Computer Vision? An Automated Visual Analysis of News Media Images on Facebook. *Social Media + Society*, *9*(3), 20563051231195582. https://doi.org/10.1177/20563051231195582

Kosmidis, S., Hobolt, S. B., Molloy, E., & Whitefield, S. (2019). Party Competition and Emotive Rhetoric. *Comparative Political Studies*, *52*(6), 811–837. https://doi.org/10.1177/0010414018797942

Koufakou, A., Nieves, E., & Peller, J. (2024). Towards a new Benchmark for Emotion Detection in NLP: A Unifying Framework of Recent Corpora. In D. Hupkes, V. Dankers, K. Batsuren, A. Kazemnejad, C. Christodoulopoulos, M. Giulianelli, & R. Cotterell (Eds.), *Proceedings of the 2nd GenBench Workshop on Generalisation (Benchmarking) in NLP* (pp. 196–206). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.genbench-1.13

Kuran, T. (1991). Now Out of Never: The Element of Surprise in the East European Revolution of 1989. *World Politics*, *44*(1), 7–48. https://doi.org/10.2307/2010422

Lambert, A. J., Eadeh, F. R., & Hanson, E. J. (2019). Anger and its consequences for judgment and behavior: Recent developments in social and political psychology. In *Advances in*

*Experimental Social Psychology* (Vol. 59, pp. 103–173). Elsevier. https://doi.org/10.1016/bs.aesp.2018.12.001
Lasswell, H. D. (2009). *Power and Personality*. Transaction Publishers. https://books.google.com/books?id=1y69mxixRYQC
Le Bon, G. (2009). *Psychology of crowds*. Sparkling Books.
Lertchoosakul, K. (2021). The white ribbon movement: High school students in the 2020 Thai youth protests. *Critical Asian Studies*, *53*(2), 206–218. https://doi.org/10.1080/14672715.2021.1883452
Lim, M. (2013). Framing Bouazizi: 'White lies', hybrid network, and collective/connective action in the 2010–11 Tunisian uprising. *Journalism*, *14*(7), 921–941. https://doi.org/10.1177/1464884913478359
Lind, F., Gruber, M., & Boomgaarden, H. G. (2017). Content Analysis by the Crowd: Assessing the Usability of Crowdsourcing for Coding Latent Constructs. *Communication Methods and Measures*, *11*(3), 191–209. https://doi.org/10.1080/19312458.2017.1317338
Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2023). *Visual Instruction Tuning* (arXiv:2304.08485). arXiv. https://doi.org/10.48550/arXiv.2304.08485
Lu, L., Tao, R., Kwon, H., Kang, J., Zhou, Y., Xin, H., Duncan, J., & McLeod, D. (2025). Visual Constructs of Conflict and Solidarity: The Role of Visual Framing on Public Perceptions and Engagement Intentions with Social Protests. *Visual Communication Quarterly*, *32*(1), 17–32. https://doi.org/10.1080/15551393.2025.2452959
Lu, Y. (2026). Performative Propaganda Engagement: How Celebrity Fans Engage with State Propaganda on Weibo. *Political Communication*, *43*(1), 99–127. https://doi.org/10.1080/10584609.2025.2584999
Lu, Y., & Peng, Y. (2024). The Mobilizing Power of Visual Media Across Stages of Social-Mediated Protests. *Political Communication*, *41*(4), 531–558. https://doi.org/10.1080/10584609.2024.2317951
Marcondes, F. S., Gala, A., Magalhães, R., Perez de Britto, F., Durães, D., & Novais, P. (2025). Using Ollama. In F. S. Marcondes, A. Gala, R. Magalhães, F. Perez de Britto, D. Durães, & P. Novais (Eds.), *Natural Language Analytics with Generative Large-Language Models: A Practical Approach with Ollama and Open-Source LLMs* (pp. 23–35). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-76631-2_3
Marcus, G. E. (2000). Emotions in Politics. *Annual Review of Political Science*, *3*(Volume 3, 2000), 221–250. https://doi.org/10.1146/annurev.polisci.3.1.221
Marcus, G. E., Neuman, W. R., & MacKuen, M. B. (2017). Measuring Emotional Response: Comparing Alternative Approaches to Measurement. *Political Science Research and Methods*, *5*(4), 733–754. https://doi.org/10.1017/psrm.2015.65
Mattoni, A., & Teune, S. (2014). Visions of Protest. A Media-Historic Perspective on Images in Social Movements. *Sociology Compass*, *8*(6), 876–887. https://doi.org/10.1111/soc4.12173
McAdam, D. (1999). *Political process and the development of Black insurgency, 1930-1970* (2nd edition). University of Chicago Press.
McCarthy, J. D., & Zald, M. N. (1977). Resource Mobilization and Social Movements: A Partial Theory. *American Journal of Sociology*, *82*(6), 1212–1241. https://doi.org/10.1086/226464
Mohammad, S., Bravo-Marquez, F., Salameh, M., & Kiritchenko, S. (2018). SemEval-2018 Task 1: Affect in Tweets. In M. Apidianaki, S. M. Mohammad, J. May, E. Shutova, S.

Bethard, & M. Carpuat (Eds.), *Proceedings of the 12th International Workshop on Semantic Evaluation* (pp. 1–17). Association for Computational Linguistics. https://doi.org/10.18653/v1/S18-1001

Molder, A. L., Lakind, A., Clemmons, Z. E., & Chen, K. (2022). Framing the Global Youth Climate Movement: A Qualitative Content Analysis of Greta Thunberg's Moral, Hopeful, and Motivational Framing on Instagram. *The International Journal of Press/Politics*, *27*(3), 668–695. https://doi.org/10.1177/19401612211055691

Morrison, D. R., & Isaac, L. W. (2012). Insurgent Images: Genre Selection and Visual Frame Amplification in IWW Cartoon Art. *Social Movement Studies*, *11*(1), 61–78. https://doi.org/10.1080/14742837.2012.640530

Neufeld, K. H. S., Starzyk, K. B., & Gaucher, D. (2019). Political Solidarity: A Theory and a Measure. *Journal of Social and Political Psychology*, *7*(2), 726–765. https://doi.org/10.5964/jspp.v7i2.1058

Neumayer, C., & Rossi, L. (2018). Images of protest in social media: Struggle over visibility and visual narratives. *New Media & Society*, *20*(11), 4293–4310. https://doi.org/10.1177/1461444818770602

Olesen, T. (2013). "We are all Khaled Said": Visual Injustice Symbols in the Egyptian Revolution, 2010–2011. In N. Doerr, A. Mattoni, & S. Teune (Eds.), *Advances in the Visual Analysis of Social Movements* (Vol. 35, pp. 3–25). Emerald Group Publishing Limited. https://doi.org/10.1108/S0163-786X(2013)0000035005

Olesen, T. (2015). *Global Injustice Symbols and Social Movements*. Palgrave Macmillan US. https://doi.org/10.1057/9781137481177

Osnabrügge, M., Hobolt, S. B., & Rodon, T. (2021). Playing to the Gallery: Emotive Rhetoric in Parliaments. *American Political Science Review*, *115*(3), 885–899. https://doi.org/10.1017/S0003055421000356

Palczewski, C. H. (2005). The Male Madonna and the Feminine Uncle Sam: Visual Argument, Icons, and Ideographs in 1909 Anti-Woman Suffrage Postcards. *Quarterly Journal of Speech*, *91*(4), 365–394. https://doi.org/10.1080/00335630500488325

Pearlman, W. (2013). Emotions and the Microfoundations of the Arab Uprisings. *Perspectives on Politics*, *11*(2), 387–409. https://doi.org/10.1017/S1537592713001072

Peng, Y., Lock, I., & Ali Salah, A. (2024). Automated Visual Analysis for the Study of Social Media Effects: Opportunities, Approaches, and Challenges. *Communication Methods and Measures*, *18*(2), 163–185. https://doi.org/10.1080/19312458.2023.2277956

Peng, Y., Lock, I., & Salah, A. A. (2024). Automated Visual Analysis for the Study of Social Media Effects: Opportunities, Approaches, and Challenges. *Communication Methods and Measures*. (world). https://www.tandfonline.com/doi/abs/10.1080/19312458.2023.2277956

Peng, Y., Lu, Y., & Shen, C. (2023). An Agenda for Studying Credibility Perceptions of Visual Misinformation. *Political Communication*, *40*(2), 225–237. https://doi.org/10.1080/10584609.2023.2175398

Popkin, S. L. (1994). *The reasoning voter: Communication and persuasion in presidential campaigns* (2nd ed). University of Chicago Press.

Prasad, I. S., & Steinert-Threlkeld, Z. C. (2025). Taken at face value: Emotion expression and protest dynamics. *Research & Politics*, *12*(3), 20531680251366095. https://doi.org/10.1177/20531680251366095

Ramsey, E. M. (2000). Inventing citizens during World War I: Suffrage cartoons in the woman citizen. *Western Journal of Communication*, *64*(2), 113–147. https://doi.org/10.1080/10570310009374668

Ransan-Cooper, H., A. Ercan, S., & Duus, S. (2018). When anger meets joy: How emotions mobilise and sustain the anti-coal seam gas movement in regional Australia. *Social Movement Studies*, *17*(6), 635–657. https://doi.org/10.1080/14742837.2018.1515624

Rossi, L., Neumayer, C., Henrichsen, J., & Beck, L. K. (2023). Measuring Violence: A Computational Analysis of Violence and Propagation of Image Tweets From Political Protest. *Social Science Computer Review*, *41*(3), 905–925. https://doi.org/10.1177/08944393211055429

Rouzegar, H., & Makrehchi, M. (2024). Enhancing Text Classification through LLM-Driven Active Learning and Human Annotation. In S. Henning & M. Stede (Eds.), *Proceedings of the 18th Linguistic Annotation Workshop (LAW-XVIII)* (pp. 98–111). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.law-1.10

Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š., & Sedlmair, M. (2018). More than Bags of Words: Sentiment Analysis with Word Embeddings. *Communication Methods and Measures*, *12*(2–3), 140–157. https://doi.org/10.1080/19312458.2018.1455817

Safaian, D., & Teune, S. (2022). Visual framing in the German movements for gay liberation and against nuclear energy. *Visual Studies*, *37*(4), 272–283. https://doi.org/10.1080/1472586X.2021.1940265

Scholz, S., Weidmann, N. B., Steinert-Threlkeld, Z. C., Keremoğlu, E., & Goldlücke, B. (2025). Improving Computer Vision Interpretability: Transparent Two-Level Classification for Complex Scenes. *Political Analysis*, *33*(2), 107–121. https://doi.org/10.1017/pan.2024.18

Shah, T. M. (2024). Emotions in Politics: A Review of Contemporary Perspectives and Trends. *International Political Science Abstracts*, *74*(1), 1–14. https://doi.org/10.1177/00208345241232769

Shahin, S., & Ng, Y. M. M. (2022). Connective action or collective inertia? Emotion, cognition, and the limits of digitally networked resistance. *Social Movement Studies*, *21*(4), 530–548. https://doi.org/10.1080/14742837.2021.1928485

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv:1409.1556). arXiv. https://doi.org/10.48550/arXiv.1409.1556

Sobolev, A., Chen, M. K., Joo, J., & Steinert-Threlkeld, Z. C. (2020). News and Geolocated Social Media Accurately Measure Protest Size Variation. *American Political Science Review*, *114*(4), 1343–1351. https://doi.org/10.1017/S0003055420000295

Song, H., Tolochko, P., Eberl, J.-M., Eisele, O., Greussing, E., Heidenreich, T., Lind, F., Galyga, S., & Boomgaarden, H. G. (2020). In Validations We Trust? The Impact of Imperfect Human Annotations as a Gold Standard on the Quality of Validation of Automated Content Analysis. *Political Communication*, *37*(4), 550–572. https://doi.org/10.1080/10584609.2020.1723752

Steinert, -Threlkeld Zachary C., Chan, A. M., & Joo, J. (2022). How State and Protester Violence Affect Protest Dynamics. *The Journal of Politics*, *84*(2), 798–813. https://doi.org/10.1086/715600

Summers-Effler, E. (2007). The emotional significance of solidarity for social movement communities: Sustaining Catholic worker community and service. In *Emotions and social movements* (pp. 135–149). Routledge. https://books.google.com/books?hl=zh-TW&lr=&id=LTOefxu2lUkC&oi=fnd&pg=PA135&dq=The+emotional+significance+of+solidarity+for+social+movement+communities&ots=CxGh0m6abi&sig=zitXjgqzh5pohoxC3Mx3mNZJwIA

Team, G., Abd, S. E., Aggarwal, V., Algayres, R., Andreev, A., Bachem, O., Ballantyne, I., Brick, C., Cărbune, V., Casbon, M., Chaturvedi, M., Cotruta, V., Coucke, A., Culliton, P., Dadashi, R., Dixon, L., Elhawaty, M., Evci, U., Farabet, C., … Zou, C. (2026). *Gemma 4 Technical Report* (arXiv:2607.02770). arXiv. https://doi.org/10.48550/arXiv.2607.02770

Teo, E., & Fu, K. (2021). A novel systematic approach of constructing protests repertoires from social media: Comparing the roles of organizational and non-organizational actors in social movement. *Journal of Computational Social Science*, *4*(2), 787–812. https://doi.org/10.1007/s42001-021-00101-3

Torres, M., & Cantú, F. (2022). Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data. *Political Analysis*, *30*(1), 113–131. https://doi.org/10.1017/pan.2021.9

Urman, A., Ho, J. C., & Katz, S. (2021). Analyzing protest mobilization on Telegram: The case of 2019 Anti-Extradition Bill movement in Hong Kong. *PLOS ONE*, *16*(10), e0256675. https://doi.org/10.1371/journal.pone.0256675

Valentino, N. A., Hutchings, V. L., & White, I. K. (2002). Cues that Matter: How Political Ads Prime Racial Attitudes During Campaigns. *American Political Science Review*, *96*(1), 75–90. https://doi.org/10.1017/S0003055402004240

van Atteveldt, W., van der Velden, M. A. C. G., & Boukes, M. (2021). The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms. *Communication Methods and Measures*, *15*(2), 121–140. https://doi.org/10.1080/19312458.2020.1869198

van Stekelenburg, J., & Klandermans, B. (2013). The social psychology of protest. *Current Sociology*, *61*(5–6), 886–905. https://doi.org/10.1177/0011392113479314

Van Troost, D., Van Stekelenburg, J., & Klandermans, B. (2013). Emotions of Protest. In N. Demertzis (Ed.), *Emotions in Politics* (pp. 186–203). Palgrave Macmillan UK. https://doi.org/10.1057/9781137025661_10

Vance, L. D., & Potash, J. S. (2022). Black Lives Matter protest art: Uncovering explicit and implicit emotions through thematic analysis. *Peace and Conflict: Journal of Peace Psychology*, *28*(1), 121–129. https://doi.org/10.1037/pac0000584

Veg, S. (2016). Creating a Textual Public Space: Slogans and Texts from Hong Kong's Umbrella Movement. *Journal of Asian Studies*, *75*(3), 673–702. https://doi.org/10.1017/S0021911816000565

Walgrave, S., Wouters, R., & Ketelaars, P. (2016). Response Problems in the Protest Survey Design: Evidence from Fifty-One Protest Events in Seven Countries*. *Mobilization: An International Quarterly*, *21*(1), 83–104. https://doi.org/10.17813/1086/671X-21-1-83

Wang, S., Jui, I. J., & Thorpe, J. (2024). Is Crowdsourcing a Puppet Show? Detecting a New Type of Fraud in Online Platforms. *Proceedings of the New Security Paradigms Workshop*, 84–95. https://doi.org/10.1145/3703465.3703472

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Webster, S. W., & Albertson, B. (2022). Emotion and Politics: Noncognitive Psychological Biases in Public Opinion. *Annual Review of Political Science*, *25*(Volume 25, 2022), 401–418. https://doi.org/10.1146/annurev-polisci-051120-105353

Wetzel, C. (2012). Envisioning Land Seizure: Diachronic Representations of the Occupation of Alcatraz Island. *American Behavioral Scientist*, *56*(2), 151–171. https://doi.org/10.1177/0002764211419354

Widmann, T. (2021). How Emotional Are Populists Really? Factors Explaining Emotional Appeals in the Communication of Political Parties. *Political Psychology*, *42*(1), 163–181. https://doi.org/10.1111/pops.12693

Wijermars, M., & Lokot, T. (2022). Is Telegram a “harbinger of freedom”? The performance, practices, and perception of platforms as political actors in authoritarian states. *Post-Soviet Affairs*, *38*(1–2), 125–145. https://doi.org/10.1080/1060586X.2022.2030645

Won, D., Steinert-Threlkeld, Z. C., & Joo, J. (2017). Protest Activity Detection and Perceived Violence Estimation from Social Media Images. *Proceedings of the 25th ACM International Conference on Multimedia, MM ’17*, 786–794. https://doi.org/10.1145/3123266.3123282

Xu, Z., Khatri, V., Dai, Y., Liu, X., Li, S., Zhang, X., & Yu, R. (2026). *Enhancing LLM-Based Data Annotation with Error Decomposition* (arXiv:2601.11920). arXiv. https://doi.org/10.48550/arXiv.2601.11920

Young, L. E. (2019). The Psychology of State Repression: Fear and Dissent Decisions in Zimbabwe. *American Political Science Review*, *113*(1), 140–155. https://doi.org/10.1017/S000305541800076X

Young, L., & Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication*, *29*(2), 205–231. https://doi.org/10.1080/10584609.2012.671234

Yuen, S., Tang, G., Lee, F. L. F., & Cheng, E. W. (2022). Surveying Spontaneous Mass Protests: Mixed-mode Sampling and Field Methods. *Sociological Methodology*, *52*(1), 75–102. https://doi.org/10.1177/00811750211071130

Zhang, H., & Leung, R. (2026). Joint Text-and-Image Clustering for Social Science Research. *Sociological Methodology*, *56*(1), 1–30. https://doi.org/10.1177/00811750251382929

Zhang, H., & Pan, J. (2019). CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media. *Sociological Methodology*, *49*(1), 1–57. https://doi.org/10.1177/0081175019860244

Zhang, H., & Peng, Y. (2024). Image Clustering: An Unsupervised Approach to Categorize Visual Data in Social Science Research. *Sociological Methods & Research*, *53*(3), 1534–1587. https://doi.org/10.1177/00491241221082603

Zhu, Y., Cheng, E. W., Shen, F., & Walker, R. M. (2022). An Eye for an Eye? An Integrated Model of Attitude Change Toward Protest Violence. *Political Communication*, *39*(4), 539–563. https://doi.org/10.1080/10584609.2022.2053915

## Appendix A Prompt for coding data from 2019 Hong Kong Protests

prompt_hk_v4 = """
Objective: Analyze the provided image from the 2019 Hong Kong Anti-Extradition Bill Protests within its historical framework (mass protests demanding democratic reforms, police accountability, and protection of civil liberties). Evaluate visual evidence for Solidarity, Anger, and Joy from the pro-protest perspective. Output strictly in JSON.

### 1. STEP-BY-STEP WORKFLOW
1. TRANSCRIBE & TRANSLATE: Locate and transcribe all salient text/slogans. Translate Cantonese into English internally.
2. DESCRIBE: Provide a concise description of the visual layout and key subjects (max 30 words).
3. EVALUATE & CODE LAST: Write detailed qualitative explanations for each emotion FIRST (max 30 words). Then, assign the binary presence code ("0" or "1") and certainty score [0-10] based strictly on your analysis.

### 2. DEFINITIONS & CUES
DEFINITIONS
- Solidarity: Positive collective emotion from shared identity, mutual aid, and hopeful commitment that sustains engagement across diverse protest factions.
- Anger: Reactive moral outrage against injustice or perceived violations of rights, typically targeting authorities to frame a clear "common enemy."
- Joy: Delight, playfulness, or relief used to counter stress, boost morale, and sustain engagement through humor and lighthearted expression.

CUES
- Solidarity: Shared memories, core protest slogans (5 demands), calls for mutual support/safety, petitions, calls from unions and narratives highlighting collective strength and international backing.
- Anger: Direct criticism of government/police, documentation of misconduct (force, arrests, abuses), and framing of systemic institutional injustice. Other targets may include pro-government citizens and businessmen.
- Joy: Satire, memes, caricatures of authority, pop culture references, and cartoonish/playful visuals used for political expression or comic relief.

### 3. STRICT RULES OF ANALYSIS
1. Holistic Evaluation: Analyze the image as a complete unit (all visuals and text together) and focus on cues salient to a human observer.
2. Diagnostic Cues: Use the visual cues listed, but remain open to other evidence.
3. Word aesthetics: Account for artistic, handwritten, non-standard, stylized typography and varied orientations when doing the word collocation
4. Image Priority: If the text message (if provided) contradicts the image, rely on the image.
5. Salience & Primary Function: While protest materials essentially are about grievances, only code anger when the dominant tone is explicit hostility, violent confrontation, or aggressive outrage.

6. Interpretation Limits: Default to "0". Only mark an emotion as present ("1") if clear, explicit evidence exists. Do not make highly abstract or speculative interpretations.
7. Salience Prioritization: Prioritize visual cues that command attention (size, contrast, position, headlines).
8. Factual-Emotional Weight: Acknowledge that factual images (e.g., tallies, casualties, news clips) can carry emotion.
9. Accurate Subject Identification: Distinguish between police, protesters, and neutral citizens.
10. Intent and Complexity: Do not assume an image is restricted to only one dominant emotion.
11. Broad Movement Context: Evaluate images considering the history of the movement and Hong Kong.

### 4. EXPECTED CODING LOGIC EXAMPLES
- Anger: "1" if showing police firing tear gas at citizens; "0" if text only lists future rally schedules without triggering clues.
- Solidarity: "1" if showing large crowds and messages of resilience; "0" if focused purely on an isolated act of police brutality.
- Joy: "1" if utilizing a playful meme mocking authority; "0" if the tone is entirely serious or informational.

### 5. OUTPUT SCHEMA
Perform the workflow internally. Output ONLY a single, raw, valid JSON object matching this structure exactly (use "0" or "1" as your decision variables, and "0" to "10" for the certainty level):

```
{
  "Image_Description": "Text description string",
  "Explanation_Solidarity": "Short explanation string",
  "Solidarity": "0",
  "Certainty_Solidarity": "0",
  "Explanation_Anger": "Short explanation string",
  "Anger": "0",
  "Certainty_Anger": "0",
  "Explanation_Joy": "Short explanation string",
  "Joy": "0",
  "Certainty_Joy": "0"
}
```

IMPORTANT: Do not include markdown code blocks (like ```json) or conversational text.
"""

**Appendix B: Prompt for coding data from Casas & William (2019)**

prompt_blm_v4 = """
Objective: Analyze the provided social media image from the Black Lives Matter (BLM) and #ShutdownA14 protests (April 2015) within its historical framework (mass protests demanding police accountability, an end to police brutality, and racial justice). Evaluate visual evidence for Sadness, Anger, and Enthusiasm from the pro-protest perspective. Output strictly in JSON.

### 1. STEP-BY-STEP WORKFLOW
1. TRANSCRIBE & TRANSLATE: Locate and transcribe all salient text/slogans. Translate non-English text internally if applicable.
2. DESCRIBE: Provide a concise description of the visual layout and key subjects (max 30 words).
3. EVALUATE & CODE LAST: Write detailed qualitative explanations for each emotion FIRST (max 30 words). Then, assign the binary presence code ("0" or "1") and certainty score [0-10] based strictly on your analysis.

### 2. DEFINITIONS & CUES
DEFINITIONS
- Sadness: Heavy grief, sympathy, or mourning stemming from tragic loss of life, police violence, and structural factors like systemic violence and negligence toward affected families and communities.
- Anger: Reactive moral outrage against injustice or perceived violations of rights, typically targeting authorities to frame a clear "common enemy.
- Enthusiasm: A collective emotion stemming from shared identity, dynamic action, hopeful commitment, and active engagement among protesters, sustaining movement energy.

CUES
- Sadness: Active memorialization (vigils, candles, spontaneous shrines, flowers), headshots/portraits/obituaries of victims, public listings of casualty names, and authentic family mourning.
- Anger: Direct criticism of law enforcement/government, documentation of police violence/misconduct (weapons, arrests, physical force), hostile signs targeting authorities, and aggressive anti-brutality rhetoric.
- Enthusiasm: Protest flyers, calls to action, hopeful/empowering language, active marching, dynamic crowds, energetic banner-holding, and active postures signaling hopeful participation in social change.

### 3. STRICT RULES OF ANALYSIS
1. Holistic Evaluation: Analyze the image as a complete unit.
2. Diagnostic Cues: Use the visual cues listed, but remain open to other evidence.
3. Image Priority: If the text message (if provided) contradicts the image, rely on the image.
4. Ecological Grounding: Ground coding on the shared grievances of the pro-protest public.
5. Interpretation Limits: Avoid overly speculative or highly abstract interpretations.
6. Salience Prioritization: Prioritize visual cues that command attention (size, contrast, position, headlines).

7. Factual-Emotional Weight: Acknowledge that factual images (e.g., tallies, casualties, news clips) can carry emotion.
8. Accurate Subject Identification: Distinguish between police, protesters, and neutral citizens.
9. Intent and Complexity: Do not assume an image is restricted to only one dominant emotion.
10. Broad Movement Context: Evaluate images considering the history of the BLM movement and Black people.

### 4. EXPECTED CODING LOGIC EXAMPLES
- Sadness: "1" if displaying a candle-lit vigil or victim portrait; "0" if the tone is purely energetic, victorious, or informational.
- Anger: "1" if showing police aggressively deploying force or direct hostility toward officers; "0" if text only lists future march logistics or rally schedules.
- Enthusiasm: "1" if showing dynamic marching crowds or empowering calls to action; "0" if focused purely on a static memorial portrait without active engagement.

### 5. OUTPUT SCHEMA
Perform the workflow internally. Output ONLY a single, raw, valid JSON object matching this structure exactly (use "0" or "1" as your decision variables, and "0" to "10" for the certainty level):

```
{
  "Image_Description": "Text description string",
  "Explanation_Sadness": "Short explanation string",
  "Sadness": "0",
  "Certainty_Sadness": "0",
  "Explanation_Anger": "Short explanation string",
  "Anger": "0",
  "Certainty_Anger": "0",
  "Explanation_Enthusiasm": "Short explanation string",
  "Enthusiasm": "0",
  "Certainty_Enthusiasm": "0"
}
```

IMPORTANT: Do not include markdown code blocks (like ```json) or conversational text.
"""

## Appendix C: Results from alternative model on the 2019 Hong Kong Protests Data

**Table A1. Alternative Model – Results from Gemma4:31B models**

| | **Runtime** | | |
|---|---|---|---|
| Plain image | 20s | | |
| With message | 21s | | |
| *Performance: Plain image* | **Anger** | **Solidarity** | **Joy** |
| Precision | 0.58 | 0.77 | 0.46 |
| Recall | 0.89 | 0.93 | 0.70 |
| F1-Score | 0.70 | 0.84 | 0.56 |
| *Performance: With Message* | **Anger** | **Solidarity** | **Joy** |
| *Precision* | 0.59 | 0.76 | 0.49 |
| *Recall* | 0.87 | 0.95 | 0.76 |
| F1-Score | 0.70 | 0.85 | 0.59 |
| *Cohen's Kappa with Qwen3-VL:32B* | **Anger** | **Solidarity** | **Joy** |
| *Plain image* | 0.70 | 0.44 | 0.48 |
| *With Message* | 0.61 | 0.47 | 0.47 |

**Appendix D: Comparison of LLMs**

**Table A2. Comparing features and performance of the three LLMs employed**

| Areas of focus | Gemma 3 | Llava | Qwen3-vl |
|---|---|---|---|
| 1. Vision encoding | **Big-picture summaries:** Shrinks every photo to a small, fixed summary size. Great at identifying general scenes and reading standard text across 140+ languages. | **Clear, front-and-center images:** Chops images into square tiles to keep details clear. Works well for clear photos with obvious subjects | **Micro-details & bad handwriting:** Excels at reading small, blurry, or tilted text on protest signs. Can pinpoint exact locations of objects in using bounding boxes. |
| 2. How smart is its reasoning? | **Solid general logic:** Good at answering straightforward questions and giving standard descriptions but misses tiny visual clues hidden in crowded backgrounds. | **Basic step-by-step logic:** Relies on its text memory to make sense of what it sees. Good for simple questions but can hallucinate details in chaotic scenes. | **Detective-level visual logic:** Uses "chain-of-thought" reasoning to think step-by-step. It looks closely at symbols, signs, and spatial relationships before reaching a conclusion. |
| 3. How fast and cheap is it to run? | **Balanced speed & high precision:** Because every image is compressed into a fixed, lightweight size, it processes massive piles of photos quickly using minimal computer memory. | **Slower & needs more computer power:** Tiling high-resolution photos into multiple pieces uses up a lot of memory fast, making large-scale image processing slower and pricier. | **Slowest:** Adjusts how hard it works depending on image complexity. Requires decent computer power by studying details of images with reasoning |
| 4. New dimension: customization & fine-tuning | **Moderate effort:** Supported by modern ai tools, but its fixed image size limits how much you can teach it to spot tiny new visual details. | **Easiest to customize:** Good for finetuning by connecting to many other open-source tools scoring system or categories. | **Advanced setup required:** Capable of specialized tasks, but demand higher computational power |
| 5. Main takeaway | **The high-speed worker:** pick this if you have **data at scale t**hat could be grossly classify or describe without fine-grained visual markers | **The customizer:** pick this if you want to finetune a model on your own custom rules using easy, beginner-friendly open-source tools. | **The deep investigator:** pick this if you need to read tiny sign text, count objects in a crowd, or analyze complex political symbols. |