

How media from the US and China describe AI development

1. Introduction

Artificial Intelligence (AI) has already become a core basic facility in the 21st century (Ittefaq et al., 2024). In aspects such as medical, business, education, manufacturing, and defense, AI produces deep impacts on people's daily lives, economic development, and national security (Ittefaq et al., 2024). Through the spread of AI systems, public discussions tend more to focus on the wider social, ethical, and geopolitical implications, but not only on technical performance (Nguyen & Hekman, 2022a). News media play a core role in these discussions because the usage of news media is not just describing facts, but building an explanation framework through choices of theme, words, and language (Nguyen & Hekman, 2022a). This framework affects people's understanding of AI (Nguyen & Hekman, 2022a).

The research about the AI news framework reveals that news actively constructs meanings about automation and datafication instead of reflecting the fact neutrally (Nguyen & Hekman, 2022a). Studies about major outlets in Western countries such as The New York Times, The Guardian, Wired, and Gizmodo indicate that the news about AI in the mid-2010s increased rapidly and gradually became more critical (Nguyen & Hekman, 2022a). The framework connects AI to issues like surveillance, data bias, cybercrime, and information disorder (Nguyen & Hekman, 2022a). Comparative work across 12 countries discovered that there is a repetitive global framework around the impact of AI on the business, economy, employment, education, national security, and ethics, showing that news media help define people's acknowledgment of AI's risk and profit balance (Ittefaq et al., 2024).

America and China are two leading AI powers internationally; they have a special position in terms of investment, research output, and strategic ambition (Nguyen & Hekman, 2022b; Ittefaq et al., 2024). However, the two countries' media systems in the political structure, ownership, and normative expectations have great differences (Ittefaq et al., 2024). For instance, American media such as The New York Times and CNN are working in a commercial and diverse environment, which could accommodate diverse ideological positions and emphasize investigative scrutiny, entrepreneurial innovation, and debates over ethics and regulation (Ittefaq et al., 2024). In contrast, the major Chinese media, including People's Daily and China Daily, are closely related to state priorities and often situate AI within narratives of national development, technological self-reliance, and long-term modernization goals (Ittefaq et al., 2024).

The comparative work that evaluates Chinese and American media outlets about AI emphasizes how media discourses reflect the institutional and geopolitical orientation (Nguyen & Hekman, 2022b). For instance, Nguyen's study of The Washington Post and the South China Morning Post shows that the media in both countries acknowledge AI's economic potential, but differ in how strongly they foreground political risks, global rivalry, and security concerns (Nguyen & Hekman, 2022b). American media are more likely to stress conflict, data risks, and democratic vulnerabilities (Nguyen

& Hekman, 2022b). Broader cross-national research similarly reports that Global North newspapers, including American major outlets, more often emphasize regulatory, ethical, and political problematization of AI, whereas outlets from China and other parts of the Global South more frequently highlight innovation, economic growth, and positive national futures (Ittefaq et al., 2024). These studies suggest that media not only simply reflect the existing AI reality, but also localize the global technology trend within specific national projects and value systems (Nguyen & Hekman, 2022b; Ittefaq et al., 2024).

Such patterns make the news language become the key research object of understanding how AI behaves in different countries' social and political constructions (Nguyen & Hekman, 2022a). Linguistic features such as metaphor choice, attribution of agency to governments, firms, or AI systems, and the degree of certainty or caution in reporting can show who is controlling AI, who is benefiting from AI, and who is bearing the risk (Nguyen & Hekman, 2022a; Frame Element–Based Model authors, 2024). Take American and Chinese media as an example, examining such language choices can illuminate how public discourses from both countries reflect and reinforce national ideologies, whether AI is viewed primarily as a competitive arena in a "new arms race," a driving force for high-quality development, or a governance challenge requiring coordinated regulation (Nguyen & Hekman, 2022b). Through corpus linguistics methods, this paper will discuss the following questions: (1) What linguistic patterns characterize U.S. and Chinese media coverage of AI?, and (2) How do U.S. and Chinese media frame and evaluate AI development? In this background, this study builds a specialized corpus of news articles from selected American and Chinese outlets from October to November 2025 (Ittefaq et al., 2024). The time frame is chosen because events like APEC are held during this period (Brookings Institution, 2025). Through the corpus built, this study systematically compares their linguistic framing and evaluative stances toward AI, aiming to show how the two leading powers communicate about the most consequential technology recently (Ittefaq et al., 2024; Nguyen & Hekman, 2022b).

2. Methodology

The corpus was specifically constructed to represent how media in China(including Hong Kong and Mainland) and the US describe the recent development of Artificial Intelligence(AI) recently .

2.1 data collection method and source

The data utilized in this report were sourced from the HKUST library's "Factiva" database. We collected news articles from both China(Mainland) and the United States, selecting three major media outlets from the US and two major media outlets from China, with a total of 100 articles per outlet to minimize bias and ensure consistency. The selected media outlets are as follows:

China:

1. China Daily -117,815 tokens
2. People's Daily -37,184 tokens

US:

1. CNN – 125,684 tokens
2. The New York Times – 111,196 tokens
3. Business Wire -56,874 tokens

Overall size : 448,753 tokens

Articles were retrieved individually via Factiva's advanced search using the query "AI",

2.2 Data collection time range

The data were collected specifically from October to November 2025. Given the rapid evolution of AI technology and news coverage, events such as the APEC meeting, which featured Xi Jinping's push for global AI cooperation, can significantly influence how media portray AI developments. Therefore, the dataset was collected over a recent and brief time frame to accurately reflect current trends in AI.

2.3 Ethical considerations

Ethical considerations were carefully observed throughout the corpus-building process. All articles were retrieved from Factiva, a commercial news aggregator accessed through an institutional subscription provided by the HKUST Library. The selected texts are published journalistic pieces that are already in the public domain and intended for wide dissemination. No copyrighted material was used beyond the scope permitted by the library's license agreement, and no private correspondence, leaked documents, or restricted-access content was incorporated. Since the corpus consists entirely of

professional news reporting and contains no personal identifiers or individual-level data, no additional anonymization was required. The dataset has been compiled and analyzed solely for academic and educational purposes as part of this assessment and will not be redistributed or used for commercial purposes.

2.4 Data cleaning procedure

Data collected from the website were downloaded in RTF format and cleaned using Python code, which included the following steps:

1. Expand contractions (e.g., I'm → I am, they're → they are)
2. Remove punctuation to ensure only meaningful text remains
3. REMOVE symbols , document identifiers, email addresses , contact information blocks from all the articles
4. Convert all text to lowercase for consistency in analysis

Once the cleaning process was complete, the data were converted to text files with UTF-8 encoding. These files were then processed using TagAnt for part-of-speech (POS) tagging, resulting in six text files, each named according to its corresponding media outlet. Additionally, A master Excel spreadsheet was maintained containing the following metadata for each of the 600 articles: article title, publication date, exact URL, publisher, word count, and a brief note on main AI-related topic. This ensured full traceability and reproducibility.

2.5 Data Analysis Method

The analysis was conducted using AntConc, a tool that facilitates corpus analysis through various methods, including keyword identification, collocation, and coordination. The first stage of analysis involved word frequency analysis (using normalized frequency for consistency), which allowed us to identify content words that reflect key themes and priorities. This was followed by a keyword analysis that compared data from China and the United States, helping to identify distinctive terms that differentiate the two countries in their descriptions of AI development. Related keywords were grouped into themes or semantic fields. Subsequently, N-gram analysis was performed by adjusting cluster sizes, enabling us to identify lexical bundles that serve rhetorical or evaluative functions and compare findings between China and the United States. These analytical steps provided a systematic and comprehensive exploration of the dataset, laying the groundwork for deeper investigation.

The next phase aimed at identifying words that frequently co-occur with selected target words to uncover typical associations, meanings, and evaluative patterns in context. Key terms extracted from the frequency and keyword lists, such as “economics” and “development,” were subjected to cluster analysis with varying cluster sizes of 3-5 words. This helped at identify recurring phrase frames or

multi-word patterns surrounding each target word, enhancing our understanding of how ideas are framed by the media. Collocation analysis was also conducted using a 5L–5R span window, employing mutual information (MI) as the statistical measure for this study to detect idiomatic or unexpected combinations. Finally, a concordance (KWIC) analysis was performed to confirm the meanings and contexts of these patterns for our findings and interpretations.

This multi-step procedure facilitated a systematic progression from broad patterns to detailed contextual interpretation, transitioning from exploratory analyses (word frequency, keyword lists, N-grams) to focused analyses (clusters, collocates, concordance lines) for interpretation and findings. The results of the analysis are presented in the following section.

3. Analysis and Findings

Based on an analysis of Chinese and American media reports on AI, which include 227,559 tokens from China and 334,872 tokens from the US, we found two main differences. The first is how the political intentions behind AI development are framed as a relationship between cooperation and competition. The second is the focus on collective versus individual agency in AI applications.

3.1 Cooperation & Competition in AI Discourse

Keywords and collocations show that Chinese and American media have distinctly different geopolitical strategies in AI development. Table 1 presents the main geopolitical keywords from the two corpora.

Table 1.1: Top Geopolitical Keywords in AI Discourse

China Corpus	Rank	Freq	NormFreq		US Corpus	Rank	Freq	NormFreq
development	18	1146	5036.057		trump	33	948	3228.036
global	28	736	3234.326		company	40	734	2499.345
cooperation	37	622	2733.357		china	63	490	1668.500
south	42	512	2249.966		president	81	409	1392.687
apec	48	458	2012.665		world	85	388	1321.179
growth	52	443	1946.748		companies	88	345	1174.760
market	53	442	1942.353		data	89	343	1167.950
system	54	440	1933.564		know	93	307	1045.366
trade	56	422	1854.464		tech	94	296	1007.910
quality	59	401	1762.180		million	98	280	953.428

*For china, “us(a)” only have about 150 frequency

The Chinese corpus features terms centered around multilateral institutions and cooperation, such as “cooperation” (2733.36 NormFreq) and “APEC” (2012.67 NormFreq). In contrast, the American corpus focuses more on political figures (“Trump” at 3228.04 NormFreq) and companies (“company” 2499.35 NormFreq). Although “China” (1668.50 NormFreq) appears less frequently, it mostly shows up in competitive contexts.

Example 1 (China):

*“...strong country will inevitably seek hegemony we champion win win **cooperation** and denouncing the winner takes all practices we advocate mutual...”*

Example 2 (China):

*“...the business centric gathering ran in parallel with the annual **APEC** economic leaders meeting. AI was the focus of this...”*

Example 3 (China):

*“...he sees great potential for the two countries to deepen **cooperation** in ai development lee*

young chan a professor at... ”

These concordance lines frame AI development as fundamentally collaborative. Example 1 emphasizes "win-win cooperation" while rejecting "winner takes all" approaches, Example 2 situates AI discussions within APEC's multilateral framework, and Example 3 highlights "potential to deepen cooperation" between nations. The frequent collocation of "cooperation" with "APEC economic leaders" (78 instances, 342.78 NormFreq) reinforces this narrative by anchoring AI discourse in formal multilateral institutions, presenting China as championing collective progress rather than pursuing unilateral gains.

In contrast, the US corpus frames China through a competitive and cautious lens.

Example 4 (US):

*“...against an increasingly assertive **China** including by expanding restrictions on China's access to American high tech like the advanced...”*

Example 5 (US):

*“...their message despite Washington's attempts to slow its gains, **China** was at the forefront of the global ai...”*

Example 6 (US):

*“...course and allow nvidia to sell the H20 will help **China's** leading AI companies buy time at a...”*

Table 1.2: Cluster result for searching word “China” in US corpus

US(china) Cluster	Rank	Freq	NormFreq
assertive china	12	11	0.011
authoritarian china	18	10	0.09
increasingly assertive china	34	8	0.09
trade imbalance china	38	4	0.004
china holds significant	44	3	0.003

Here, “China” appears as a competitor and often collocates with words related to threats or assertiveness, such as “authoritarian China” (keyness = 73.88), “assertive China” (keyness = 69.55), and “trade imbalance” This language pattern constructs China as a competitor that needs to be watched and balanced.

Table 1.3: N-gram result for China and US corpus

Chain Type	Rank	Freq	NormFreq		US Type	Rank	Freq	NormFreq
quality productive forces	14	89	391.118		president donald trump	18	55	187.286
apec economic leaders	20	78	342.777		business financial desk	26	49	166.855
economic leaders meeting	22	71	312.737		financial desk section	31	44	149.829
asia pacific region	24	62	272.464		cheering applause spanberger	58	29	98.751
explainers faqs analyses	39	50	219.729		leader xi jinping	74	26	88.535

In the collocation analysis, three of the top five most frequent three-word phrases in the Chinese corpus relate to multilateral cooperation: "APEC economic leaders" (78 Freq), "economic leaders meeting" (71 Freq), and "Asia Pacific region" (62 Freq). The American corpus, by contrast, foregrounds political leadership and corporate actors through phrases like "President Donald Trump" (55 Freq), "business financial desk" (49 Freq), and "financial desk section" (44 Freq).

The data above reveal how different geopolitical contexts shape the discourse around artificial intelligence. Chinese media portray AI as an area of multilateral cooperation and shared development, which aligns with China's broader foreign policy and its focus on regional partnerships. In contrast, American media present AI through the lens of national security and competition, reflecting the U.S. concern over technological rivalry and maintaining AI leadership. This fundamental difference in framing leads to distinct narratives of cooperation versus competition in Chinese and American media, which runs throughout the language used to discuss AI.

3.2 Collective & Individual Agency in AI Application

Besides geopolitical differences, the two corpora also show clear contrasts in how agency and innovation in AI applications are portrayed. Table 2 compares the top collocates related to agency and action in AI contexts.

Table 2.1: Agency Collocates in AI Discourse

China Corpus	Frequency		US Corpus	Frequency
development	1146		Trump	948
economic	674		company	734
growth	443		companies	345
promote	301		million	280
governance	238		OpenAi	216

In the Chinese corpus, high-frequency words like “development” (1146 instances), “economic” (674 instances), and “governance” (238 instances) reflect collective, state-coordinated agency. Terms such as “promote” (301 instances) indicate that AI development is institutionally guided and closely tied to the country’s economic growth.

Example 7 (China):

*“...focus should be on improving the quality of **development** high **quality development** does not negate the importance of quantitative growth...”*

Example 8 (China):

*“...set the momentous goals on various fronts including pursuing high **quality development** fostering technological self reliance and enhancing the chinese...”*

Table2.2:Collocates for China corpus searching keyword “quality development”

Collocate	Rank	Freq(N)	FreqLR	FreqL	FreqR	Likelihood
standard	2	720	15	3	12	84.097
economy	3	3060	19	4	15	62.352
advance	4	1480	12	6	6	45.308

Table2.3 : Keyword and Cluster (searching word “system”) for China

Keyword Type (China)	Rank	Freq	NormFreq
system	54	440	1933.564
Cluster Chain (system)	Freq		
industrial system	25		

innovation system	9		
-------------------	---	--	--

These examples show that AI applications are described as top-down and carefully planned. The phrase “quality development” often appears alongside words like “standard,” “economy,” and “advance,” suggesting that AI serves national goals of economic transformation and sustainable growth. Notably, the word “system” appears frequently (440 times, normalized frequency 1933.56) and often collocates with phrases like “innovation system” and “industrial system.” This indicates that the concept of AI is driven by a framework of planned systems rather than market forces.

Table 2.4: N-Gram and Cluster (searched word:”development”) for China

N-Gram Type (China)	Rank	Freq	NormFreq
quality productive forces	14	89	391.118
Cluster(development)	Rank	Freq	
integrated development	13	30	

The cluster further confirms this finding. Phrases like “quality productive forces” (89 Freq), “integrated development” (30 Freq), and “national development” reflect the characteristic of centralized planning, positioning AI as a collective national development goal.

In stark contrast, the US corpus attributes agency primarily to individual companies and entrepreneurs.

Example 9 (US):

*“...earlier this year that would have brought it closer to **OpenAi**’s chatgpt and **Google**’s gemini rivals like Google...”*

Example 10 (US):

*“...is no accident **huang** has become a central figure in **trump**’s push for the us to dominate ai and...”*

Table 2.5:Selected Keyword for US corpus compared to China corpus

Keyword (US) Type	Rank	NormFreq q_Tar	Keyness (Likelihood)		Keyword (US) Type	Rank	Freq	NormFreq
meta	18	735.502	247.917		company	40	734	2499.345
openai	23	820.629	214.022		companies	88	345	1174.760
google	27	711.666	179.798					
oscar	34	493.740	166.411					
musk	38	473.309	149.304					
cooper	46	394.992	133.124					

*left side are for individual company/person , while the right side are the general term(e.g. company) for them

In the American corpus, sentences with company names like Meta and Google as the subject are very common, with these companies often playing the main role in reports. The high frequency of words like “company” (2499.35 NormFreq) and “companies” (1174.760 NormFreq), along with company names such as “OpenAI” (820 NormFreq_Tar), together create an image of an AI market driven by

entrepreneurship and competitive business. Individual leaders also appear frequently, with names like “Trump” (3228.036 NormFreq) and CEOs such as Musk, Cooper, and Oscar, emphasizing individual agency rather than institutional frameworks.

This stands in sharp contrast to China’s collective approach. The Chinese corpus reflects coordinated planning, while the American corpus shows a market-driven environment where companies “race” and “invest billions” to advance AI technology. These represent two fundamentally different development models: a state-led approach versus an entrepreneur-driven capitalist model.

3.3 Contrasting AI Narratives

Table 3: table for summary for analysis

Dimension	China Corpus	US Corpus
Geopolitical Frame	Cooperation	Competition
Top Keywords	development, global, cooperation, APEC	Trump, company, China (as competitor)
Actor Representation	State, institutions	Companies, individual leaders
Development Model	Quality, systemic	Market-driven
Evaluative Prosod	Positive (cooperation, win-win)	Mixed (innovation + concern)

The analyses in Sections 3.1 and 3.2 reveal consistent ideological portrayals. Chinese media describe AI through language emphasizing multilateral cooperation (APEC, win-win) and collective development coordinated by the state (quality development, systems). In contrast, American media depict AI from a competitive and individualistic perspective, portraying China as a strategic competitor while viewing companies as the main drivers of innovation.

These differences reflect underlying political-economic structures. China’s discourse aligns with centralized planning and a diplomatic focus on regional cooperation. The U.S. discourse, on the other hand, is shaped by capitalist markets and geopolitical competition. More importantly, the two also show clear contrasts in semantic prosody. Chinese language frames AI cooperation positively (“win-win,” “advance”), while U.S. language adopts a competitive and cautious tone in describing China’s AI efforts (“authoritarian,” “trade imbalance”), highlighting their differing ideological stances.

These findings show that media in different regions assign very different meanings and descriptions to AI development. The language used is not neutral but serves different purposes. For example, in the Chinese context, AI is seen as a tool for collective progress, while in the American context, AI is viewed as a strategic area that requires competition. These differences in reporting lead to varied public understandings of AI’s role at both the national and international levels.

4. Discussion and Conclusion

Discussion of Findings

When we examine how China and the US talk about AI, we see two fundamentally different approaches that reflect where each country stands in the global technology landscape. These contrasting viewpoints are shaping how AI will evolve in both nations.

China's position as a fast-follower has naturally led it to embrace international partnerships. We see this reflected in frequent references to collaborative platforms like APEC and mentions of other countries in Chinese discourse. Through regional frameworks like APEC—which brings together 21 economies to promote sustainable growth and trade—China is working to strengthen its AI capabilities through partnership rather than going it alone.

Meanwhile, American media coverage tends to frame China's AI progress in negative terms. This suggests that the US, holding its position as the technology leader, views China's rapid advancement with some degree of concern or skepticism about the direction other nations are taking in this space.

The ways AI gets deployed in each country tell us a lot about how public activities align with government priorities. Chinese media consistently connects AI applications to official policy language, particularly terms like "quality development." This phrase, which Xi Jinping introduced in 2017, represents a deliberate pivot from prioritizing speed to emphasizing sustainability, efficiency, and balance in economic growth. Within this framework, AI becomes a mechanism for updating existing processes to support these long-term goals.

The American discussion looks quite different. US sources regularly spotlight individual companies and prominent figures driving AI initiatives, reflecting the wide range of perspectives and projects coming from private actors. These patterns make sense given the fundamental differences in how the two economies operate.

Our analysis generally confirms what we expected to find about national attitudes toward AI and the role of geopolitical factors. The results align well with existing scholarship, including Nguyen and Hekman's work on how geopolitical positioning and strategic goals shape the relationship between AI and different economic systems. One surprise, though, was that we didn't find as much evidence of widespread public anxiety about AI as we had anticipated.

Implications and Limitations

These findings give us some clues about where AI is headed in both countries. We can reasonably expect China to continue channeling its AI efforts toward "quality development" objectives, while the US will likely keep nurturing a diverse ecosystem of industry applications and cutting-edge products, with policy evolving to match. The relationship between these two powers—whether it leans more cooperative or competitive on AI matters—is something we can continue to track.

Understanding how these global leaders communicate about AI helps us build a fuller picture of where the technology is going worldwide, making it easier to navigate this rapidly changing landscape.

That said, we need to be realistic about the study's limitations. Given how quickly AI technology evolves, these findings may only hold true for the recent past. A breakthrough technology could completely shift how AI gets applied. We should also acknowledge that the media sources we analyzed may carry their own biases, which could mean our results don't perfectly capture reality on the ground.

Concluding Remarks

Ultimately, attitudes toward AI development and application in China and the US are shaped by both external pressures and internal dynamics. The context surrounding media coverage of AI is deeply influenced by broader social, ethical, and geopolitical considerations.

Looking ahead, researchers could gain additional insights by analyzing news coverage over longer timeframes, building on what we've learned here to develop a more complete understanding of AI's trajectory.

5. References

Nguyen, D., & Hekman, E. (2025). The news framing of artificial intelligence: A critical exploration of how media discourses make sense of automation. **AI & Society*, 40*(2), 1179–1193.
<https://doi.org/10.1007/s00146-022-01511-1>

Nguyen, D., & Hekman, E. (2022b). A "new arms race"? Framing China and the U.S.A. in A.I. news reporting: A comparative analysis of the **Washington Post** and **South China Morning Post**. **Journalism*, 23*(11), 2399–2417. <https://doi.org/10.1177/20594364221078626>

Ittefaq, M., Xia, Q., & Zhang, X. (2024). Global news media coverage of artificial intelligence (AI): A comparative analysis of frames, sentiments, and topics. **Telecommunications Policy*, 48*(9), Article 102923. <https://www.sciencedirect.com/science/article/abs/pii/S0736585324001278>

Brookings Institution. (2025, October 28). **How will AI influence US–China relations in the next 5 years?**
<https://www.brookings.edu/articles/how-will-ai-influence-us-china-relations-in-the-next-5-years/>

Liu, Y., & Wang, Z. (2024). Detecting misinformation through framing theory: The frame element-based model. arXiv. <https://doi.org/10.48550/arXiv.2402.15525>

Full text of Chinese President Xi Jinping's written speech at APEC CEO Summi
https://english.www.gov.cn/news/202510/31/content_WS6904a6cac6d00ca5f9a0737c.html

Trump says rare earths deal and tariff cut agreed with China after high-stakes talks
<https://edition.cnn.com/2025/10/29/asia/us-china-trump-xi-meet-south-korea-intl-hnk>

Nvidia Chips Are Front and Center in China at Crucial Moment in the A.I. Race
<https://www.nytimes.com/2025/07/31/technology/china-ai-nvidia-chips.html>

What is "High-quality Development"? | Understand China's Policies
<https://www.ourchinastory.com/en/11871/What-is->

6. Individual Reflections

Choi Wing Yan, Amy

My primary contributions involved laying the project's groundwork through comprehensive background research and drafting the introduction and conclusion for both our presentation and final report. I was also responsible for the meticulous task of finding and verifying all academic citations.

Through this process, I developed practical corpus linguistics skills, learning to identify linguistic framing by deconstructing methodological approaches in prior research. This enabled me to recognize subtle agency attribution in news texts and understand the necessity of balanced corpora for meaningful cross-national comparison.

The work presented significant challenges, particularly technical issues with AntConc software on my personal computer, which frequently became unresponsive. Academically, locating specific essays that directly compared US and Chinese AI media coverage proved exceptionally difficult and time-consuming.

Ultimately, this project profoundly demonstrated the power of data-driven analysis to uncover hidden ideological patterns in media discourse, moving beyond mere surface description. I gained valuable insight into collaborative empirical research and now feel equipped to construct and utilize specialized corpora for future linguistic investigations, solidifying my analytical approach to language and media.

TAM, Tsz Wun Charlie

In this project, my main responsibilities were designing the corpus methodology and writing the analysis and findings chapters (sections 3.1 and 3.2). Thomas and I worked together to collect data, clean it by removing redundant information like author names and dates, and maintain detailed metadata logs. I took the lead in writing the core analysis, transforming quantitative results (keywords, collocations, n-grams) into meaningful qualitative interpretations.

Through this research, I learned how to design a balanced bilingual corpus, ensuring comparability between the Chinese corpus and American corpus in scale, timeframe, and representativeness. I mastered using AntConc for multi-level analysis, from keyword extraction to collocation analysis and KWIC concordancing. More importantly, I learned to connect quantitative data with qualitative interpretation. For instance, when I found high frequencies of "cooperation" and "APEC" in the Chinese corpus, I not only presented the numbers but also explained how these words construct narratives of multilateralism and "win-win" cooperation through specific contextual examples.

The biggest challenge was clearly presenting complex patterns in abstract geopolitical frameworks and showing relationships between data points. To overcome this, I collaborated closely with Louis to design clear table structures (Tables 1.1, 1.2, 2.1) and carefully selected concise, representative contextual examples. I also learned to write flowing paragraphs connecting different tables and examples, making the analysis section coherent and accessible to non-specialist readers.

This project taught me that corpus linguistics can reveal ideological patterns hidden in everyday language use. China's emphasis on "integrated development" versus America's focus on individual companies like OpenAI reflects fundamentally different political-economic structures. This data-driven approach to examining media discourse has strengthened my critical thinking skills for future academic research.

Chow Sheung Yau Thomas

During this project, I was primarily responsible for the data collection component, with a specific focus on China-related sources, as well as for drafting the methodology section of the report.

Through this assessment, I gained valuable experience in building a corpus from scratch. This included collecting data from websites using the “Instant Data Scraper” browser extension and writing Python scripts to clean the dataset by removing redundant or irrelevant information such as author names, publication dates, and duplicate article titles. Additionally, I developed skills in conducting systematic corpus analysis using AntConc, including extracting and comparing keywords across two datasets, adjusting cluster sizes and concordance settings, and employing KWIC (Key Word in Context) lines to substantiate our findings. These analytical steps enabled us to derive meaningful conclusions from the data.

The project also presented several challenges. In the data collection phase, we observed that the most readily available sources for U.S.-related coverage were not mainstream American media outlets, which yielded significantly fewer articles. This raised a methodological dilemma: whether to prioritize data accessibility or representational accuracy by focusing on mainstream outlets. Ultimately, we opted for the latter approach, extending the time span for U.S. sources to achieve comparable corpus size with the China-related dataset.

Furthermore, the analytical results deviated considerably from our initial hypotheses. We had anticipated prominent themes related to AI posing threats to human jobs or causing unemployment; however, keywords associated with these concepts appeared with relatively low frequency and keyness in both corpora. Consequently, we shifted our focus to alternative patterns and insights emerging from the data, which required substantial adjustment of our original research expectations.

This experience strengthened my proficiency in corpus compilation, data cleaning, and corpus linguistics tools while highlighting the importance of flexibility when empirical evidence challenges preconceived assumptions.

Fung Yat Long Louis

My specific role in the project is analysis. I am responsible for generating tables of words, keywords, N-gram, cluster and collocates. Those tables are used to discuss findings or insights we identify from the corpora. I put those tables and cloudword pictures into a word file.

I learned how to use antconc to create a corpus, filter out function words, sort the list by keyness. By this kind of analysis, I have a more clear understanding of word, keyword, n-gram and cluster. I know the meanings and differences of those concepts used in natural language processing. Moreover, I learned how to adjust the setting in Antconc to get what I want, for example using the MI to search for collocates.

One biggest challenge I have faced is choosing words to search in clusters and collocates. Some tables are hard to conclude with few meaningful words. I am afraid that I am too objective and cannot show the meaningful difference between corpora. The insight or findings will be limited if I chose the inappropriate word. After I ask the opinions from teammates, we come up with some search words. It is a valuable experience to share thoughts and come up with a conclusion after discussion.

I gained some analysing skills from this project. I also learned how to analyse from a given dataset or even given topic to some actual insights. By breaking the whole analysis into parts like background research, methodology (collect and clean data), create informative tables using platforms like antconc. It is a comprehensive lesson that I have the chance to take part in every section and learn from others.

In addition, I found this project benefits my science major skills. This project strengthens my analysis skill as a data science student. I learned how to analyze words this time instead of some scientific index. The similarity of them are trying to understand or find something from a large given data. It is a special and meaningful experience for me.

7. Division of Work

Member	main responsibility	Section in the Report	Other Duties
TAM, Tsz Wun Charlie	Corpus compilation (US related data),Findings	Analysis and Findings	Check of methodology parts presentation: corpus design and methodology
Chow Sheung Yau Thomas	Corpus compilation(China related data),Findings	Methodology	To check progress of the written report presentation: Key findings and analysis
Choi Wing Yan	Background research, Findings	Introduction	Final check on the report Presentation: Key findings and analysis
Fung Yat Long Louis	analysis(keywords/clu ster/N-gram...),Findin gs	Discussion and Conclusion	arange meetings for discussion presentation: Introduction /Discussion and takeaway