Text-to-Video Models and National Security Work: Exploring Potential, Identifying Risks, and Formulating Policies

Haitao Pan[1], Zefeng Wang[2,*], Hongchang Zhou[2], Jean-Marie Nianga[3]
[1]Anding College, Huzhou University, Huzhou, China
[2]School of Information Engineering, Huzhou University, Huzhou, China
[3]Sino-Congolaise pour le Développement (Fondation), Brazaville, Congo Republic
**Corresponding Author.*

Abstract: With the rapid development of artificial intelligence technology, Text-to-Video models, which represent the forefront of current video generation technology, will face numerous opportunities and challenges in the future. This paper combines a large number of research reports and scientific experience to introduce the pros and cons of Text-to-Video models for national security work and put forward relevant suggestions. Text-to-Video models have significant value in protecting national security and maintaining social unity. By generating high-quality video content, Text-to-Video models can effectively disseminate shared social values, display the country's history, culture, and social development achievements, enhance national cultural soft power and international discourse power, and enhance the cohesion and influence of protecting national security and maintaining social unity. However, the misuse of Text-to-Video model technology and the spread of false information also pose new challenges to protecting national security and maintaining social unity. Therefore, when using Text-to-Video models, it is necessary to emphasize the importance of their reasonable application and ensure that technological applications comply with laws regulations and social ethics, and conform to shared social values and serve the country's long-term development and social harmony and stability.

Keywords: Text-to-Video; Opportunities; Challenges; National Security; Strategy

1. Introduction

Text-to-Video models, particularly Sora, represent the forefront of current video generation technology. These models, primarily based on deep learning frameworks such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and the recently emerged Diffusion Models, have demonstrated the ability to generate high-quality video content. Notably, Diffusion Models have gained attention due to their advantages in generating high-resolution, high-quality images and videos [1]. A prominent feature of the Sora model is its ability to generate videos based on text descriptions. It includes two Diffusion Models: a text-based video generation model (T2V) and an image-to-video generation model (I2V). The T2V model synthesizes videos based on given text inputs, while the I2V model adds image inputs to generate videos that strictly adhere to the content, structure, and style of the provided reference image [1]. The development of these technologies has not only driven technological advancements in video generation but also explored potential directions for future applications.

Internationally, with the rapid development of artificial intelligence technology, video generation models, particularly text-based video generation systems like Imagen Video, have achieved high-definition video generation capabilities. These systems generate high-definition videos based on text prompts through a combination of a base video generation model and a series of interleaved spatial and temporal video super-resolution models. Imagen Video's success demonstrates the high controllability of video generation technology and its understanding of the world, including the ability to generate diverse videos and text animations in various artistic styles and 3D object comprehension [2]. In certain large countries, such as China, video generation models have also gained

widespread attention and application due to technological advancements and market demand. Chinese technology companies and research institutions have made significant progress in deep learning and video generation. The domestic market demand for video content continues to grow, especially in education, entertainment, and advertising, providing ample space for the application of Text-to-Video models.

In the new era, the focus and application of technological innovation in protecting national security and maintaining social unity have become particularly important, especially in cutting-edge technology fields like Text-to-Video models. The development of these technologies has not only revolutionized information dissemination but also provided new perspectives and methods for protecting national security and maintaining social unity. The application of Text-to-Video models can play a unique role in spreading common values [3], enhancing national cultural soft power, and strengthening international discourse power. It also provides technical support for high-quality economic development and modern social governance. By generating high-quality, highly realistic video content, Text-to-Video models can provide a new means of communication for the work of protecting national security and maintaining social unity. This technology can automatically generate videos based on text descriptions, allowing complex concepts and ideas to be presented to the public in a more intuitive and vivid way. This greatly improves the efficiency and acceptance of information dissemination. For instance, when promoting common social values, Text-to-Video models can be used to create video content with strong visual impact and appeal, allowing these values to resonate more deeply with the public.

Furthermore, Text-to-Video models demonstrate immense potential in enhancing national cultural soft power and international discourse power. Through this technology, high-quality videos showcasing a country's traditional culture, historical heritage, and modern development achievements can be created and disseminated internationally. This content can help people around the world gain a more comprehensive and in-depth understanding of the country. Not only does this contribute to enhancing the country's cultural influence internationally but also builds bridges for international exchange and cultural exchange.

The application of Text-to-Video models is equally significant in achieving high-quality economic development [4]. This technology can be used to produce corporate promotional videos [5], product demonstration videos, and the like. It helps enterprises effectively showcase their strengths and product features, enhance brand image, and ultimately promote economic development. Additionally, Text-to-Video models can be applied in vocational education and training. By generating videos that simulate real-world operation scenarios, they provide more intuitive and practical learning materials for vocational skills training, improving training effectiveness.

In terms of modernizing social governance, Text-to-Video models offer an effective tool for information dissemination and public education. Governments and social organizations can utilize this technology to produce videos on public safety education, health literacy, and legal awareness promotion [6]. This allows them to convey important information to the public in a more engaging manner, raising public safety and legal awareness, and promoting social harmony and stability.

Therefore, as a cutting-edge technological innovation, Text-to-Video models demonstrate broad application prospects and far-reaching influence in the new era's work of protecting national security and maintaining social unity. By deeply exploring and applying this technology, significant achievements can be made in spreading common values, enhancing national cultural soft power, strengthening international discourse power, promoting high-quality economic development, and modernizing social governance. This will contribute to national development and the progress of human society.

2. Opportunities and Challenges of Text-to-Video Models

2.1 New Opportunities in Protecting National Security and Maintaining Social Unity.

In the context of the new era, Text-to-Video models present novel opportunities for the work of protecting national security and

maintaining social unity, particularly in disseminating shared social values and national culture, and enhancing national cultural soft power and international discourse power. The application of these cutting-edge technologies can not only facilitate effective information dissemination but also deepen education on shared social values, promote positive cultural construction, and guide students and the public towards positive values and ethical notions.

Text-to-Video models, through their high level of innovation and interactivity, offer new pathways for disseminating shared social values and various cultures. These models can automatically generate high-quality video content based on text descriptions [7], allowing complex concepts and ideas to be presented to the public in a more intuitive and vivid manner. For instance, by creating animated films or short videos that delve into the essence of social common values, not only can students' and the public's understanding and recognition of these values be enhanced but their sense of national pride and mission can also be ignited. Furthermore, utilizing Text-to-Video models to showcase the rich historical and cultural heritage, modern development achievements, and the lifestyles of people from different countries can effectively strengthen national cultural confidence and promote the inheritance and development of excellent traditional cultures of various nations.

Text-to-Video models hold immense potential for promoting social harmony, cultural vibrancy, and global engagement. By addressing the challenges and harnessing the power of these technologies, we can effectively disseminate shared values, promote national cultures, strengthen international discourse power, and guide students and the public towards positive values and ethical notions. As these models continue to evolve, their contribution to a more harmonious, culturally vibrant, and globally engaged society will only grow.

However, the application of Text-to-Video models also faces challenges [8]. Ensuring that the generated video content accurately reflects the shared social values and the true essence of various cultures, avoiding misunderstandings and deviations, is an issue that requires careful consideration. Additionally, effectively utilizing this technology to enhance national cultural soft power and international discourse power requires in-depth research and practical exploration, taking into account the national conditions and international situation of each country.

2.2 Challenges Faced.

The potential social division and value conflicts caused by technological abuse. As a powerful content generation tool, the Text-to-Video models can generate highly realistic video content based on text descriptions. If used to produce and disseminate false information, it may mislead the public and undermine the foundation of social trust, exacerbating social conflicts and divisions. For example, generating content with biased and discriminatory characteristics may deepen the divide between different groups and exacerbate social conflicts, undermining social harmony and stability [9]. In addition, technological abuse may also lead to value conflicts. In the context of globalization, the exchange of different cultures and values is increasingly frequent. If the Text-to-Video video model is used to disseminate specific ideologies or values while ignoring cultural diversity and the inclusiveness of values, it may trigger value conflicts and affect international relations and social diversity and harmony.

The potential threat to social stability and national image from the spread of false information and harmful content [10]. Another challenge we face with the Text-to-Video models is the spread of false information and harmful content, which poses a potential threat to social stability and national image. Since the generated video content is extremely realistic, audiences with insufficient discernment skills may have difficulty identifying its authenticity and are thus susceptible to false information. This not only misleads the public's understanding of important social events, but also may be used for political manipulation, undermining election fairness, interfering with the internal affairs of other countries, and posing a threat to democratic systems and the international order. Meanwhile, the spread of harmful content, such as violence and pornography, can have a negative impact on teenagers and undermine social morals, affect the country's image and cultural security. Therefore, how to effectively regulate the use of Text-to-Video models to prevent the spread of false information and harmful content is a

major challenge facing us today.

3. National Response Strategy

To address these challenges, it is necessary to strengthen the research on technological ethics, improve laws and regulations, establish sound regulatory mechanisms, while enhancing public media literacy education to enhance society's ability to identify and prevent false information. This will ensure the healthy development of deepfake video models and serve the modernization construction.

3.1 Strengthening Ideological and Political Education.

In light of the challenges posed by Text-to-Video models, particularly in terms of technology misuse and the spread of misinformation, strengthening ideological and political education has emerged as a crucial countermeasure. This necessitates not only innovation in the content and form of ideological and political education but also the enhancement of ideological security in cyberspace to ensure the effective dissemination of shared social values and social stability.

Leveraging Generative Text-to-Video models to Innovate Ideological and Political Education Content and Forms. Text-to-Video models, as an advanced technological tool, hold immense value for ideological and political education. This technology enables the generation of highly realistic and engaging video content on demand, opening up new possibilities for innovating ideological and political education content and forms. For instance, Text-to-Video models can be employed to create animations or short films on topics such as world history, culture, and shared social values [11]. These materials can be used to convey ideological and political education content to the public, particularly the younger generation, in a more vivid and intuitive manner. Such content is more readily accepted and understood, effectively enhancing the appeal and impact of ideological and political education. Furthermore, Text-to-Video models, can be utilized to simulate social practice activities. Through virtual reality technology, students can immerse themselves in virtual environments to experience social practices [12], fostering their sense of social responsibility and historical mission. This innovative educational approach allows students to learn and absorb shared social values through interactive experiences, leading to a deeper understanding and identification with these values.

Strengthen ideological security in cyberspace. With the development of information technology, cyberspace has become a new frontline in ideological struggle. The emergence of Text-to-Video video models has brought greater challenges to cyberspace ideological security [13]. Therefore, strengthening cyberspace ideological security has become an important strategy to respond to the Text-to-Video video challenge. This requires starting from the following aspects: Firstly, strengthen the supervision and management of online content to prevent the spread of false information and harmful content. The government and relevant departments should establish a sound online content supervision mechanism and step up efforts to combat false information and harmful content to ensure a clear cyberspace [14]. Secondly, enhance the public's, especially young people's, online literacy and improve their ability to identify false information. This can be achieved through school education, public media, etc. to popularize online literacy education and educate the public on how to distinguish the authenticity of online information and how to properly handle various information on the internet. Finally, actively build a positive and healthy online culture and promote the online dissemination of social common values. By producing and promoting high-quality content, the cyberspace can be filled up and online public opinion can be guided, thus creating a positive and upward-looking online environment [15].

3.2 Strengthening Laws, Regulations, and Technological Supervision.

In response to the challenges presented by Text-to-Video models, strengthening laws, regulations, and technological supervision is a crucial strategy. This necessitates not only the development of specific laws and regulations tailored to Text-to-Video models, but also the establishment of robust technological review and content moderation mechanisms to prevent technology misuse.

The rapid advancement of artificial intelligence technologies like Text-to-Video

models necessitates the development of specific laws and regulations tailored to address the emerging challenges and potential risks associated with these novel technologies [16]. Existing legal frameworks may struggle to fully encompass the diverse issues arising from these advancements, making the formulation of specialized laws and regulations for Text-to-Video models an urgent priority. These regulations should clearly define the legitimate scope of application for Text-to-Video models, establishing fundamental principles and standards for the creation and dissemination of content generated using these models. The objective is to ensure that technological applications do not infringe upon individual and collective rights or cause harm to social order and the public interest. For instance, the laws and regulations should explicitly prohibit the use of Text-to-Video models to create and distribute false information, content that infringes on copyright or portrait or any material that could incite social panic, hatred, or violence [17]. Furthermore, for individuals and organizations utilizing Text-to-Video models for content creation and dissemination, the laws and regulations should mandate the fulfillment of corresponding information disclosure obligations, ensuring that the public can discern the source and authenticity of the content.

Establish and improve technical review and content supervision mechanisms. In addition to strengthening laws and regulations, implementing robust technological review and content moderation mechanisms is another critical measure to prevent the misuse of Text-to-Video models technology. This encompasses, but is not limited to, establishing specialized technological review systems that conduct both pre-release screening and post-release supervision of content generated using Text-to-Video models, ensuring that all content adheres to legal requirements and societal ethical norms. Technological review mechanisms should leverage cutting-edge information technologies [18], such as artificial intelligence recognition techniques, to perform automated content reviews, thereby enhancing review efficiency and accuracy. Additionally, public reporting and feedback mechanisms should be established to encourage public participation in content moderation, fostering a favorable environment for social co-governance. Moreover, stringent accountability mechanisms, including but not limited to economic penalties, industry bans, and public apologies, should be implemented for individuals and organizations that violate laws, regulations, and review standards, effectively deterring potential violators.

3.3 Strengthening Technological Support for Safeguarding National Security and Maintaining Social Unity.

In the face of the challenges posed by Text-to-Video models, strengthening the technological underpinnings of efforts to safeguard national security and maintain social unity has emerged as a crucial strategy. This encompasses not only utilizing technological innovation to promote unity and exchange among diverse social groups but also employing Text-to-Video models to enhance cultural connections and emotional bonds with overseas compatriots. The implementation of these strategies aims to harness the power of technology to reinforce the cohesion and influence of protecting national security and maintaining social unity, thereby contributing to a harmonious and stable social environment.

Harnessing Technological Innovation to Foster Unity and Exchange among Diverse Social Groups. Technological innovation, particularly the application of emerging technologies such as Text-to-Video models, offers new avenues for fostering unity and exchange among diverse social groups [19]. Text-to-Video models can generate highly realistic video content that can be used to showcase the cultural characteristics, lifestyles, and values of different groups, thereby enhancing mutual understanding and respect. For instance, Text-to-Video models can be utilized to create a series of short films introducing the cultures of various nations and ethnicities, which can then be disseminated widely through online platforms, enabling more people to learn about the rich tapestry of global cultures and promoting ethnic unity. Moreover, Text-to-Video models can also be employed to simulate multiple perspectives on social hot-button issues, helping different social groups to comprehend each other's positions and aspirations, thereby reducing misunderstandings and conflicts. This approach not only strengthens societal

inclusivity and harmony but also ignites public enthusiasm for participation in social governance, fostering collective progress.

Leveraging Text-to-Video models to Enhance Cultural Connections and Emotional Bonds among Diverse Groups. Using China as an example, Text-to-Video models offer a novel approach to strengthening cultural ties and emotional bonds with overseas Chinese and compatriots in Hong Kong, Macao, and Taiwan. By producing high-quality cultural dissemination videos, the essence of traditional Chinese festivals, significant social development achievements, and the nation's future development blueprint can be effectively conveyed, enhancing their sense of identity and belonging to their ancestral homeland. For instance, Text-to-Video models can be employed to create videos introducing the customs of traditional festivals like Chinese Spring Festival and Mid-Autumn Festival, which can be shared with overseas Chinese and compatriots in Hong Kong, Macao, and Taiwan through internet platforms, enabling them to experience the vibrant festive atmosphere and patriotic sentiment even while residing overseas. Additionally, videos showcasing China's technological advancements, cultural prosperity, and social harmony can be produced to foster their confidence and pride in China's development. These measures will not only strengthen cultural ties and emotional bonds with overseas Chinese and compatriots in Hong Kong, Macao, and Taiwan but also promote their understanding and support for the nation's development, contributing to the shared belief of all Chinese people, both within and beyond the country's borders, in safeguarding national security, maintaining social unity, and realizing the Chinese Dream.

3.4 Promoting International Cooperation and Exchange.

In Response to the Challenges of Text-to-Video models, Fostering International Cooperation and Exchange Emerges as a Crucial Strategy. As Text-to-Video models continue to revolutionize the realm of content creation and dissemination, the potential challenges associated with this technology have also come to the forefront. To effectively address these concerns and ensure the responsible development and application of Text-to-Video models, fostering international cooperation and exchange has emerged as a critical strategy.

Actively Engaging in International Forums and Organizations to Promote the Development of International Rules and Standards. With the rapid advancement of emerging technologies like Text-to-Video models, the international community urgently needs to establish a set of consensus and rules to guide the development and application of these technologies [20]. In this process, actively participating in international forums and organizations and playing a constructive role is crucial for promoting the development of international rules and standards. This can not only help form an international consensus on the application of emerging technologies but also promote the healthy development of technologies and prevent the negative impacts of technology misuse. For instance, initiatives can be proposed in international organizations such as UNESCO and the ITU, calling on the international community to jointly focus on the ethical and social implications of technologies like Text-to-Video models and jointly explore and formulate relevant international rules and standards. Through these international platforms, mutual understanding and trust can be enhanced between countries, and international rules and standards can be jointly formulated that both promote technological innovation and ensure that the application of technology complies with ethical and legal requirements.

Sharing Experiences with Other Countries to Collectively Address the Challenges of Technological Advancements. In the endeavor to foster international cooperation and exchange, sharing experiences with other nations to jointly address the challenges posed by technological advancements emerges as another crucial strategy. By establishing bilateral or multilateral cooperation mechanisms, we can facilitate the exchange of technological information and experiences, enabling collective exploration of effective strategies to mitigate the potential challenges presented by emerging technologies like Text-to-Video models.

4. Conclusion

Text-to-Video models, as an advanced artificial intelligence technology, hold

immense value in the realm of protecting national security and maintaining social unity. By generating high-quality, highly realistic video content, Text-to-Video models can effectively disseminate common social values, showcase the history, culture, and social development achievements of various countries, and enhance national cultural soft power and international discourse power. This not only helps to enhance the public's sense of identity and belonging to common social values but also promotes communication and understanding among different social groups both domestically and internationally, strengthening the cohesion and influence of safeguarding national security and maintaining social stability.

However, the misuse of Text-to-Video models and the spread of misinformation also pose new challenges to protecting national security and maintaining social unity. Therefore, when utilizing Text-to-Video models, it is crucial to emphasize the importance of their rational application, ensuring that the technology application adheres to laws, regulations, and social ethics, aligns with common social values, and serves the country's long-term development and social harmony and stability.

References

[1] Haoxin Chen, Menghan Xia, Yin-Yin He, et al. “VideoCrafter1: Open Diffusion Models for High-Quality Video Generation.” arXiv preprint arXiv: 2310. 19512, 2023.

[2] Jonathan Ho, William Chan, Chitwan Saharia, et al. “Imagen Video: High Definition Video Generation with Diffusion Models.” arXiv preprint arXiv: 2210. 02303, 2022.

[3] Mijat Kustudic, Gustave Florentin Nkoulou Mvondo, et al. “A Hero or A Killer? Overview of Opportunities, Challenges, and Implications of the Text-to-Video Model SORA”.

[4] Joseph Cho, Fachrina Dewi Puspitasari, Sheng Zheng, Jingyao Zheng, Lik-Hang Lee, Tae-Ho Kim, Choong Seon Hong, Chaoning Zhang. “Sora as an AGI World Model? A Complete Survey on Text-to-Video Generation.” arXiv preprint arXiv:2403.05131 [cs.AI] (or arXiv:2403.05131v1 [cs.AI] for this version).

[5] Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, Hong Lin, et al. “AI-Generated Content (AIGC): A Survey” arXiv preprint arXiv: 2304. 06632 [cs.AI] (or arXiv: 2304. 06632v1 [cs.AI] for this version).

[6] Adetayo, A.J., Enamudu, A.I., Lawal, F.M. and Odunewu, A.O. "From text to video with AI: the rise and potential of Sora in education and libraries", Library Hi Tech News, Vol. ahead-of-print No. ahead-of-print, 2024.

[7] Yaosi Hu, Chong Luo, Zhenzhong Chen , “Make It Move: Controllable Image-to-Video Generation With Text Descriptions” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 18219-18228 arXiv:2112.02815[cs.CV].

[8] Jesus Perez-Martin, Benjamin Bustos, Silvio Jamil F. Guimarães, Ivan Sipiran, Jorge Pérez & Grethel Coello Said, “A comprehensive review of the video-to-text problem”.

[9] Tunboson Oyewale Oladoyinbo, Samuel Oladiipo Olabanji, Oluwaseun Oladeji Olaniyi, et al, “Exploring the Challenges of Artificial Intelligence in Data Integrity and its Influence on Social Dynamics” Asian Journal of Advanced Research and Reports, Volume 18, Issue 2, Page 1-23, 2024.

[10] Jacobeen, S. “The Potential Impact of Video Manipulation and Fraudulent Simulation Technology on Political Stability.” In: Kosal, M.E. (eds) Proliferation of Weapons- and Dual-Use Technologies. Advanced Sciences and Technologies for Security Applications. Springer, Cham, 2021.

[11] Zhengliang Liu, Yiwei Li, Qian Cao, Junwen Chen, Tianze Yang, Zihao Wu, John Hale, John Gibbs, Khaled Rasheed, Ninghao Liu, Gengchen Mai, Tianming Liu, “Transformation vs Tradition: Artificial General Intelligence (AGI) for Arts and Humanities” arXiv:2310.19626 [cs.AI] (or arXiv:2310.19626v1 [cs.AI] for this version).

[12] Lau, K. W., & Lee, P. Y. The use of virtual reality for creating unusual environmental stimulation to motivate students to explore creative ideas.

Interactive Learning Environments, 2015; 23(1), 3–18.

[13]Ronald J. Deibert, Rafal Rohozinski, Risking Security, “Policies and Paradoxes of Cyberspace Security”, International Political Sociology, Volume 4, Issue 1, March 2010; 15–32.

[14]Chinese Academy of Cyberspace Studies. Rule of Law Construction in Cyberspace. In: China Internet Development Report 2020. Springer, Singapore, 2023.

[15]Kui, Y. Governance Change and Political Identity in the Internet Age. Social Sciences in China, 2019: 40(4), 129–147.

[16]Yixin Liu, Kai Zhang, Yuan Li, Zhiling Yan, Chujie Gao, Ruoxi Chen, Zhengqing Yuan, Yue Huang, Hanchi Sun, Jianfeng Gao, Lifang He, Lichao Sun, “Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models” arXiv:2402.17177[cs.CV](or arXiv:2402.17177v3 [cs.CV] for this version).

[17]Van der Sloot, Bart, “Regulating the Synthetic Society ----Generative AI, Legal Questions, and Societal Challenges”.

[18]Tan Ching Ng Sie Yee Lau, Morteza Ghobakhloo, et al, “The Application of Industry 4.0 Technological Constituents for Sustainable Manufacturing: A Content-Centric Review” Sustainability 2022; 14(7), 4327.

[19]Yao Lyu, He Zhang, Shuo Niu, Jie Cai, “A Preliminary Exploration of YouTubers' Use of Generative-AI in Content Creation” arXiv:2403.06039 [cs.HC](or arXiv:2403.06039v1 [cs.HC] for this version).

[20]Mittelstadt, B. Principles alone cannot guarantee ethical AI. Nat Mach Intell, 2019; 1, 501–507.

