Text-to-video Large-language Models in Cybercrime Litigations:

Theoretical Applications & Ethics

Nicholas S. Caudill

American Military University

INTL451: Cybercrime

Instructor: Brian Burnett

July 28, 2024

nicholas.caudill@mycampus.apus.edu

Introduction

Imagine converting descriptions of technical actions within computer network topologies that are relevant to cybercrime investigations into a form that jurors or policymakers can easily visualize like watching a movie or well-animated presentation. Technology like OpenAI's Sora or Google's Veo are the cutting-edge of text-to-video large-language models (T2V LLMs). Logged computer network data could be used as input for T2V LLMs to generate plausible videos showcasing a suspected criminal's computer screen from their own point of view. Our thesis is to describe the application of T2V LLMs within cybercrime litigations to increase the likelihood of a fair trial by increasing the competency of jurors.

For a simple example, say a network activity log shows host A pinged host B on a device running Windows 10 and includes the respective timestamps and ports used. A T2V LLM could reconstruct a video of a Windows operating system user typing within the terminal followed by consequences of those actions within a visualization of the computer network, followed by showcasing how the victim was allegedly harmed. This would convert the confusing flood of information picked up by computer network logs and create an aesthetically pleasing presentation that jurors could visualize and possibly lead to jurors better comprehending defense and prosecutorial arguments.

Per Cunningham (2020, 209), using techniques like mutual Transport Layer Security (mTLS) and device validation ensures that packets transmitted within computer networks can be traced and validated. These techniques help forensic data become 'beyond a reasonable doubt' and provide telemetry analysis on network communications which would create a solid labeled dataset prior to being integrated with T2V LLMs (Cornell Law School, 2020). One example of collecting forensic evidence might include making a copy of the suspect's computer's hard drive, hashing those copies, and using a write blocker to ensure that the integrity of the collected data abides by provisions granted by warrants (Graham & Smith, 2020, 116-120).

We presume T2V LLMs have been trained on a large corpus of YouTube tutorials, so certain keywords logged and made within a command-line interface (CLI) should prompt the model to associate the text of these commands to a video showing the perspective of a graphical user interface (GUI) within a given operating system. Combining these technologies and methods may reduce the knowledge barrier when describing cyberspace concerns to policymakers or jurors as most prefer evidence in a GUI form rather than a CLI form. It is currently impractical to reverse engineer or convert CLI data into a GUI visual representation without the adoption of T2V LLMs. Being able to visualize, in movie format, the alleged damage or harm caused by suspects via creating vivid topological and 3-dimensional movies of network pipelines may help illustrate causes and effects allegedly performed by suspects.

Technical Overview of One Application

Videos are a series of frames separated by equal units of time. Picture frames, photos, or images can be expressed simply as matrices where each element's value corresponds to an 8-bit gray-scale pixel value between 0 (black) to 256 (white) (Gamli *et al.* 2018; Leo, 2023). We elect 'black and white' videos as these are computationally cheaper for analysis than colored videos. A video can be expressed as a multidimensional array or tensor or 'layers of matrices,' which allows for the application of properties from linear algebra and other computational techniques.

LLMs in general are widely known to occasionally hallucinate their output (Vipula *et al.*, 2023). Our method outlined here suggests that the negative effects of LLM hallucinations on their applicability in investigations, which are aimed to be used within courtrooms to augment a solid base of evidence to enhance juror competency, might be diminished by comparing a great amount of independently produced videos or tensors generated by T2V LLMs using prosecutorial or defensive evidence.

Denote the tensor (i.e. a multi-dimensional matrix representation of the T2V LLM's output) as $[A]$ by using the logged behavior of the cybercriminal suspect and other computer forensics data legally acquired as input for the T2V LLM. We prompt the agent to reconstruct what the activity would look like from the cybercriminal suspect's perspective at the moment of allegedly committing the crime. We might call $[A]$ the suspect's action tensor. We generate many variants of $[A]$ denoted as $[A]_n$ to diminish the negative effects of LLM hallucinations later on.

Similarly, we ask a T2V LLM to produce another tensor based on the legal wording of current legislation. A T2V LLM is prompted to construct a video showcasing an example of proper computer use for comparison purposes during litigations. We call this the law's action tensor $[L]$. Likewise, we mitigate the effect of LLM hallucinations by generating a series of video-example tensors denoted as $[L]_n$. A tensor is just a mathematical way of describing a static video.

Next, we place both series of law and action tensors along each axis of a probability adjacency matrix. By calculating the average value after summing all elements of the adjacency matrix, we can measure how much the suspect's behavior deviates or aligns with the law. For example, the jury might find any value above 0.50 to be persuasive in justifying a conviction.

The dimensions of each tensor would be $244\ x\ 244\ x\ DT$. 244 pixels is the standard size for training computer vision models based on historical reasons (Howard & Gugger, 2020). DT is the sampling rate or frame rate that turns our series of images into a continuous video by showing frames at a speed in which motion blurs and the deception of continuous motion occurs for observers. For example, if we have the T2V LLM produce a video with a duration (D) of 1-minute or 60 seconds, a framerate (T) of 60 frames-per-second, and with a resolution of $244\ x\ 244$ pixels, then the dimensions of the video tensor would be $244\ x\ 244\ x\ 3600$ (i.e. 3600 layers of 244 x 244 matrices). Thus, videos generated by T2V LLMs depicting proper and perceived legal behavior can be mathematically analyzed and compared by formalizing the video as 3,600 different matrices that form a rectangular-prism-like tensor chopped into smaller cubic units and comparing the likeness between each synaptic connection.

Next, we repurpose HyperNEAT or similar algorithms that were previously used for "detecting anomalous network conditions" to compare two tensors, say $[L]_1$ and $[A]_1$, and then assign a probability value between 0 and 1 that will be placed inside an adjacency matrix for investigative analysis (Zhukabayeva *et al.*, 2024). For Figure 1, summing each of the sixteen elements equals 8.78, and $\frac{8.78}{16}$ equals an average percentage value of 0.54875 or 55%. Thus, the jury might use this analysis to conclude that the suspect was acting more or less with or against the law. If values greater than zero means the suspect's alleged actions deviate further away from the law as the value approaches 1.0, then a 0.55 average value might entice a jury to find the suspect guilty of cybercrime.

Figure 1 showcases how videos generated by T2V LLMs, expressed as computer-friendly tensors, might be compared and evaluated by neural network models like generative adversarial networks (GANs) (Yinka-Banjo &Ugot, 2019). The idea of a ‘probability’ adjacency matrix may be a novel contribution by the author of this paper in supposing that it is possible to infer the existential likelihood of edges between the vertices of a graph (Poole, 2015, 236). Figure 1 is an example of comparing eight total T2V LLM-generated videos in tensor format, but the amount of tensors can be arbitrarily increased given the amount of computational resources available.

Figure 1

Probability Adjacency Matrix Comparing T2V LLM-Generated Tensors

$$\begin{array}{c|cccc|}
[A]_4 & .41 & .50 & .98 & .41 \\
[A]_3 & .31 & .94 & .11 & .55 \\
[A]_2 & .63 & .16 & .81 & .91 \\
[A]_1 & .54 & .12 & .77 & .63 \\
\hline
 & [L]_1 & [L]_2 & [L]_3 & [L]_4
\end{array}$$

Theoretical Overview of Other Applications

While Mogavi *et al.* (2024) highlight concerns that T2V LLMs like OpenAI's Sora or Google's Veo have the potential for malicious use within cyberspace like propagating deepfakes and disinformation, this paper adopts a more optimistic approach in using T2V LLMs as a tool to persuade jurors that the suspect of a cybercrime is innocent or guilty by simply using the investigative evidence as input to generate a movie for the courtroom to watch. A lawyer defending a client might prompt a T2V LLM to generate a movie showcasing the high moral and social status of the defendant using an autobiographical book written by the defendant themselves to be used as textual input for the T2V LLM. T2V LLMs are a revolutionary and important technology because, prior to the invention of this technology, the amount of time it would take to develop a high quality movie might interfere with the rights of a defendant for a speedy trial, especially if the defendant does not have the financial luxuries to finance the whole production themselves (Cornell Law School, n.d.).

The use of movies, videos, or films within courtrooms has been documented. For example, Lederer (2001) writes "A video deposition, unlike a typed transcript, allows a trial jury to consider the demeanor of a witness while testifying." The use of video over text within courtrooms allows jurors to obtain a wider perspective of the accused. T2V LLMs might usher in a new era of judicial science in experimenting what are the most effective and persuasive ways to display evidence and defenses within courtroom litigations.

Ethics

Fei-Fei Li, a top artificial intelligence (AI) researcher and chief scientist at Google Cloud made several important ethical arguments in her book that are relevant to the use of T2V LLMs within courtrooms (Li, 2023). One lesson Li (2023) learned when integrating prediction models within hospitals to ensure and remind healthcare workers to wash their hands was that the nurses and doctors were not afraid of the proposed application in itself, but were more afraid of the potential for the application to become a stepping stone towards more dystopian uses. Li (2023) also highlights a growing concern in the centralization of AI applications in describing how academia is no longer a place to achieve results in AI research as the massive corporate resources are necessary to train models on large datasets; cloud computing is not always a solution especially for real-time or time sensitive AI applications where latency becomes killer. For example, OpenAI's ChatGPT uses more than 30,000 Nvidia Graphics Processing Units (GPUs) where each costs around $10,000 (Liu, 2023). Thus, if T2V LLM adoption becomes prominent within court rooms, the undemocratic nature of the current state of AI products and research could result in unfavorable or biased parties where corporations might have a vested interest in rejecting a defendant from using their model to help their legal case. There is also the cybersecurity concern in how to prompt third-party T2V LLMs without leaking the suspect's sensitive data where current academic work is being done on encrypting LLM human-machine interfaces (Lin *et al.*, 2024).

Conclusion

Our thesis was to describe the application of T2V LLMs within cybercrime litigations to increase the likelihood of a fair trial by increasing the competency of jurors. This paper is an attempt to respond to Prayudi & SN (2015, 5) who wrote “There must be a good interface so that the data generated by digital investigators can be understood by judges and other law enforcement agencies in accordance with the applicable law.” We described how T2V LLMs may be a solution by generating videos that showcase legal uses of computers which could then be contrasted by videos generated which illustrate the alleged criminal behavior of a suspect. We described how legislative statute, YouTube videos, defendant autobiographies, and logged computer network activity may be used as textual input for T2V LLMs. We described a technical mathematical treatise on how likeness between generated videos might be compared to inform jurors on how the suspect’s behavior might have deviated more or less from the law. We ended by reviewing pragmatic ethical concerns surrounding the use of T2V LLMs within courtroom litigations.

References

Cornell Law School. (n.d.). Overview of right to a speedy trial. LII / Legal Information Institute.https://www.law.cornell.edu/constitution-conan/amendment-6/overview-of-right-to-a-speedy-trial.

Cornell Law School. (2020). Beyond a reasonable doubt. LII / Legal Information Institute. https://www.law.cornell.edu/wex/beyond_a_reasonable_doubt.

Cunningham, C. (2020). *Cyber warfare - truth, tactics, and strategies: Strategic concepts and truths to help you and your organization survive on the battleground of cyber warfare.* Packt Publishing.

Gamli, Ö. F., Eraslan, Z., & Akben, S. B. (2018). Determination of the protective effects of olive leaf extracts on microbiological and physicochemical properties of pepper paste using the image processing methods. *Journal of Food Process Engineering*, 41(7), e12861. https://doi.org/10.1111/jfpe.12861.

Graham, Roderick S. & S. K. Smith. (2020). *Cybercrime and digital deviance.* New York, NY: Routledge. https://doi.org/10.4324/9781003283256.

Howard, J., & S. Gugger. (2020). *Deep learning for coders with fastai and PyTorch: AI applications without a PhD.* O'reilly Media. https://course.fast.ai/Resources/book.html.

Lederer, F. (2001). The effect of courtroom technologies on and in appellate proceedings and courtrooms. *William & Mary Law School Faculty Publications.* https://scholarship.law.wm.edu/facpubs/1654.

Leo, C. W. (2023). What is the gray scale of colors? Color with Leo. https://www.colorwithleo.com/what-is-the-gray-scale-of-colors/.

Li, F.-F. (2023). *The Worlds I See.* Flatiron Books: A Moment of Lift Book.

Lin, G., W. Hua, & Y. Zhang. (2024). PromptCrypt: Prompt encryption for secure communication with large language models. *ArXiv (Cornell University).* https://doi.org/10.48550/arxiv.2402.05868.

Liu, Z. (2023). ChatGPT will command more than 30,000 Nvidia GPUs: Report. Tom's Hardware. https://www.tomshardware.com/news/chatgpt-nvidia-30000-gpus.

Poole, D. C. (2015). *Linear algebra: A modern introduction.* Cengage Learning.

Prayudi, Y., & SN, A. (2015). Digital chain of custody: State of the art. *International Journal of Computer Applications*, 114(5), 1–9. https://doi.org/10.5120/19971-1856.

Vipula, R., A. Sheth, P., & A. Das. (2023). A survey of hallucination in large foundation models. *ArXiv (Cornell University).* https://doi.org/10.48550/arxiv.2309.05922.

Yinka-Banjo, C., & Ugot, O. A. (2019). A review of generative adversarial networks and its application in cybersecurity. *Artificial Intelligence Review.* https://doi.org/10.1007/s10462-019-09717-4.

Zhukabayeva, T., A. Adamova, K.Ven-Tsen, Z. Nurlan, Y. Mardenov, & N. Karabayev (2024). Network attack detection using NeuroEvolution of augmenting topologies (NEAT) algorithm. JOIV: *International Journal on Informatics Visualization*, 8(1), 387–394. https://doi.org/10.62527/joiv.8.1.2220.